NANYANG TECHNOLOGICAL UNIVERSITY

Division of Economics, EGC
School of Humanities and Social Sciences
Nanyang Technological University
14 Nanyang Drive
Singapore 637332

Uncovering The Deep Roots of Conflict

James B. Ang and Satyendra K. Gupta

2017

EGC Report No: 2017/02

HSS-04-90A
Tel: +65 6790-6073
Email: D-EGC@ntu.edu.sg

UNCOVERING THE DEEP ROOTS OF CONFLICT

James B. Ang* Satyendra Kumar Gupta†

Abstract

This research establishes that the emergence and persistence of intrastate conflict incidence since 1960 are influenced by regional agro-ecological factors captured by the extent of variation in crop yield potential. Our results based on cross-country and grid-level analysis indicate that higher potential crop yield variability within a country that is exogenous to both human intervention and regional culture increases the likelihood of intrastate conflict. Our findings are robust to the inclusion of various geographical, institutional, and potentially confounding economic development correlates.

JEL Code: E02; F50; Q10

Keywords: Intrastate conflict, crop yield, agronomy

*Department of Economics, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332. Phone +65 65927534, Fax +65 67946303, E-mail: James.Ang@ntu.edu.sg.

†Department of Economics, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332. E-mail: satyendr001@e.ntu.edu.sg

Acknowledgements: Helpful comments and feedback received from Christopher Paik, Giovanni Ko, and seminar participants of Nanyang Technological University and conference participants at the 2016 WEAI conference are greatly appreciated. We gratefully acknowledge Oded Galor and Ömer Özak for sharing their crop yield data. James Ang acknowledges financial support from the NTU Start-up Grant and Singapore Ministry of Education Academic Research Fund Tier 2.

1 Introduction

Intrastate conflict is a profound reality in many parts of the world. It is so phenomenally evident that nearly half of the countries in the world have experienced at least one incidence of it over the last few decades. The resulting loss of human capital and undesirable outcomes associated with the geographical spread of violence have attracted the attention of many economists, social scientists, and psychologists to understand the roots of civil conflicts.

The literature suggests two possible conditions that would give rise to the occurrence of a conflict: (1) when the opportunity cost of engaging in conflict is less than that of peaceful cooperation or negotiation; and (2) when there is increasing return from violence. The first condition emphasizes that slow economic growth and low per capita income are notable signs of low opportunity cost, and people facing such circumstances are more likely to participate in violence (Collier and Sambanis, 2002; Blattman and Miguel, 2010). Slower growth and per capita income can, in turn, be linked to lower agricultural productivity, which is often induced by variation in rainfall, temperature, and soil quality. In particular, adverse climate conditions may impact agricultural productivity, which reduces the opportunity cost of initiating and engaging in conflict and civil wars (see, e.g., Collier and Hoeffler, 1998; Collier and Hoeffler, 2004). The second condition suggests that it is possible for the winning group to extract rents during or after the conflict. For example, Deininger 2003 finds that the presence of cash crops, such as coffee plant, increases conflicts during the Ugandan civil war when the winning group expropriated rents from the agricultural produce. In a similar vein, Angrist and Kugler 2008 show that higher price of cocoa is positively associated with the occurrence of armed conflicts in Colombia.

Recent studies have taken the above arguments further by providing evidence on how natural conditions play an important role in the onset of conflict and ongoing conflict incidence in a region. Natural rainfall was, and still is, in many parts of the developing world, the dominant source of water supply for agriculture in the absence of efficient irrigation systems. Thus, several studies have exploited the association between variation in rainfall and agricultural productivity to explain conflicts and rebellions in cross-country and within-country analysis. For example, Miguel et al. (2004) provide evidence showing that extreme weather conditions are strongly associated with the occurrence and incidence of conflict. They find a positive relationship between extreme rainfall variation and the frequency of rebel and communal conflict in Africa. Chaney (2013) shows that significant deviation above or below the optimal flood level of the Nile triggered political instability in historical Egypt. Brückner (2010) also shows that droughts (periods of low rainfall) can reduce the probability of intrastate civil conflicts through population reduction in Sub-Saharan countries.

Other studies have shown that temperature and the occurrence of conflict are strongly related. In particular, Dell et al. (2012) find that the annual variation in temperature is strongly linked to standards of living, where for every one degree Celsius increase in temperature per capita income is predicted to fall by 1.4 percent in poor countries. Similarly, Schlenker and Lobell (2010) find that variation in temperature and precipitation during yield time negatively affects agriculture productivity in African countries. Analyses using data from historical temperate and cold locations conducted by Zhang et al. (2007) show that anomalously cold events during cold epochs are associated with episodes of political instability in dynastic China and feudal Europe. In a similar vein, Hsiang et al. (2011) find that 21% of all the civil conflicts since 1950 could be attributable to the warmer and dryer weather in the continental tropics caused by El Nino/Southern Oscillation (ENSO).

This study extends the strand of literature that explores the influence of climate variables on long-term development. We argue that agricultural productivity potential has played a significant role in shaping the direction of human history following the agricultural transition. Focusing on agroecology as the main explanation for conflict is especially relevant since our physical environment (such as temperature, humidity, water supply, soil conditions, sunlight and terrain) may shape the development of human behavior, social norms, and culture, which subsequently determine the existence of peace or conflicts in a society (see, e.g., Anderson, 1987; Van de Vliert et al., 1999; Nunn and Puga, 2012; Alesina et al., 2013; Jimenez-Ayora and Ulubaşoğlu, 2015). The agroecology of a region, until the advent of modern agricultural techniques such as fertilizers, mechanized farming tools and irrigation facilities, determined the level of calorie one could get in any region.

Building on above insight, we hypothesize that there would be frequent and persistent attempts to conquer and control territories that produce high caloric yield. Higher regional variation in caloric yield should therefore, increase the frequency of these conflicts to control the most valuable resource in pre-industrial societies. Specifically, we demonstrate that the variation in agroecology, measured as the standard deviation in crop yield (in millions of kilocalories per hectare per year) within a country based on the data of Galor and Özak (2016), has a significant positive effect on conflict incidence.[1] The findings are robust to alternative measures of conflict incidence – such as, the intensity of conflict, annual frequency of conflict incidence – and the control of various geographic characteristics and other indicators of early development.

This study is also related to the literature that studies the relationship between civil war and horizontal inequality (see, e.g., Gurr, 1970; Cederman et al., 2011). Different groups of people are endowed with different levels of agricultural productivity that may have resulted in unequal economic prosperity during the preindustrial period. Consequently, a feeling of deprivation among the groups endowed with less productive agricultural land may lead to the occurrence of civil wars.

Our work is closely related to that of Galor and Özak (2015), who provide evidence that, in comparison with another commonly used land productivity measure (i.e., agricultural suitability),[2] mean crop yield is more strongly correlated with pre-industrial population density, which determines the path of subsequent development.[3] Galor and Özak (2016) also provide new evidence showing

[1]The crop yield index is a composite measure that takes into account of the effect of temperature and moisture on the growth of 48 crops. Potential yield estimation for each crop in each agro-climatic grid (with cell size of approximately 100 km2) is provided by Global Agro-Ecological Zones (GAEZ) project of Food and Agriculture Organization (FAO). For detailed description of the crop yield data, see Galor and Özak (2015).

[2]The agricultural suitability indices have been used in a number of research on long-term development, including but are not limited to Michalopoulos (2012), Ashraf and Galor (2013), Ang (2015), Ashraf and Michalopoulos (2015) and Olsson and Paik (2016).

[3]Other related studies by Depetris-Chauvin and Özak (2015, 2016) show that population density affects the division of labor and economic specialization, which in turn result in the formation of institutions and states.

that pre-industrial agro-climatic characteristics that were favorable to agricultural productivity, proxied by agricultural crop yield, induced learning processes that have had a persistent effect on the formation of long-term oriented cultures in the modern day.

Our study is also complementary to the contribution of Nunn and Qian (2014), who investigated the implication of humanitarian food aid on conflict. They argue that US food aid is often stolen by armed rebels or appropriated by the military of the ruling government, and hence an increase in aid may exacerbate civil conflict in recipient countries. Using a novel identification strategy, their results provide strong support for the view that higher US wheat aid tends to increase the incidence and duration of civil conflicts. In a similar vein, Arbatli et al. (2015) demonstrate that genetic diversity has a robust and persistent effect on the frequency, incidence, and onset of civil conflicts. However, unlike Nunn and Qian (2014) and Arbatli et al. (2015) who examine the roles of contemporary wheat aid and genetic diversity, respectively, the main departure of our study is that we focus on whether the variability of pre-industrial agricultural productivity can potentially account for the persistent occurrence of conflict.

While the current paper focuses on explaining civil conflicts since 1960, in the sensitivity checks we show that our hypothesis is equally valid for non-state conflicts where none of the group participating in the conflicts is associated with the state. All these groups, whether state or non-state, operate under the same institutional framework and try to control the same natural endowment to empower their own group. Our results also show that historic conflict since 1500 can be explained by the variation in potential crop yield. This result is unsurprising given that geographical conditions have not changed much over time. However, examining the link between interstate wars and their differences in potential crop yield variability is beyond the scope of the current study.

The rest of the paper is organized as follows. Section 2 sets out the approach and describes the data. Section 3 presents and discusses the empirical results. Some robustness tests are also performed. Section 4 presents empirical results for repeated cross-section data. In section 5, we divide the world into small grids of 2.5 by 2.5 decimal degrees to address potential concerns of endogeneity. We find that the incidence of conflicts in each grid cell over the period 1989-2015 can be explained by variation in potential (average) crop yield in respective grid cells after accounting for other geographical characteristics. The last section concludes the paper.

2 Empirical Approach and Data

2.1 Regression model

We estimate the following regression model to evaluate the influence of variation in crop yield on conflict incidence:

$$Conflict_i = \alpha + \beta sd\_cropyield_i + \gamma avg\_cropyield_i + \delta Controls_i + \epsilon_i \tag{1}$$

where $Conflict_i$ is the average number of 5-year sub-periods that experience any conflict in country i, $sd\_cropyield_i$ is the log of variation in potential (average) crop yield across cells (measured in millions of kilocalories per hectare per year) in country i, $avg\_cropyield_i$ is the log of mean potential (average) crop yield across cells in country i, $Controls_i$ is a vector of control variables included in the regressions to minimize the possibility of getting spurious estimates, and ϵ_i is the error term. We expect β to be positive and statistically significant. Our baseline sample used in the cross-country analysis consists of data for up to 135 countries.

2.2 Data

2.2.1 Outcome measures

The dependent variables are constructed using data over the period 1960-2014. We follow the common practice in the existing literature by focusing on the post-1960 period, given that most of the former European colonies in Sub-Saharan Africa, Middle East, South and East Asia gained independence since 1960. This is to reduce the possibility that the results are confounded by colonial powers influencing the domestic decision in these countries. Furthermore, as it is the common practice in the existing literature on conflict, we divide the time period of study; 1960 to 2014, into eleven non-overlapping 5-year sub-periods.

Our main source of data for ongoing conflict incidence and the onset of new conflict are the Armed Conflict Dataset, version 4-2015 and Conflict Onset and Incidence Dataset (Pettersson and Wallensteen, 2015), both of which are taken from Uppsala Conflict Data Program (UCDP). UCDP defines conflict as *"a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths."* As a consequence, any conflict episode where the number of battle deaths is less than the prescribed threshold is not recorded in this dataset. These two datasets provide an annual binary indicator of whether the threshold of 25 battle-related deaths has been exceeded in that year.

Conflict incidence. We use intrastate conflict incidence variables from the Armed Conflict Dataset. Our main outcome variable is *confinc* that takes value 1 when there is at least one conflict of any intensity in any year within a 5-year sub-period; otherwise, it is 0. This is the main outcome variable used in our baseline estimations. This is a standard approach used in the literature (see, e.g., the work of Esteban et al., 2012). Additionally, we use two other binary variables that are defined using different thresholds. Specifically, *confinc_mid* takes a value of 1 when there is at least one conflict episode having battle-related deaths of less than one thousand in a single year, but having cumulative battle-related deaths exceeding one thousand since the beginning of the episode in any year within a 5-year sub-period; otherwise it is 0. *confinc_war* takes a value of 1 when there is at least one conflict episode having battle-related deaths exceeding one thousand in a single year in any year within the 5-year sub-period; otherwise it is 0. For the simple cross-country analysis, we use the sum of the respective conflict incidence outcome variables across the 11 sub-periods. Thus, the outcome variables have values ranging between 0 and 11.

The onset of conflict. We also use intrastate conflict onset variables from Conflict Onset and Incidence Dataset. One issue with the data is that the onset of conflict could reflect an increase in battle-related deaths from less than 25 to more than 25 in an ongoing conflict, rather than the onset of a new conflict. Despite this limitation, Conflict Onset and Incidence Dataset is the best available data to measure the onset of the new conflict. To overcome the impact of measurement errors in recording the onset of a new conflict, we create two new variables using the Conflict Onset and Incidence Dataset. These variables, labeled as *onset2* and *onset5*, take a value of 1 if a new episode of conflict starts after 2 and 5 years of 'peace', respectively, in each of the five-year periods. We take the aggregate values over the eleven periods and use them as our onset variables in the simple cross-sectional analysis.

Ethnic conflict. As all civil conflicts are generally not ethnic-driven , using the UCDP dataset, Wimmer et al. (2009) have created a unique conflict dataset that focuses only on ethnic conflicts. They coded the aim of each rebel organization and indicated whether they pursued ethno-nationalist aims and recruited along ethnic lines. They identify "ethnic" as aiming to achieve ethnonational self-determination, a more favorable ethnic balance-of-power in government, ethno-regional autonomy, the end of ethnic and racial discrimination, language and other cultural rights, and the likes. In ethnic wars, armed organizations also recruit fighters predominantly among their leaders' ethnic group and forge alliances on the basis of ethnic similarity. We use the latest version of their dataset covering the period from 1960 to 2010 to obtain a measure of new ethnic conflict.

Non-state conflict. All the alternative definitions and different thresholds of minimum battle-related deaths that have been considered so far are based on the data where one side of the conflict is the government or the state forces, fighting against one or more internal armed opposition groups. These conflict incidences are more severe and prevalent in the conflict data. However, government or state forces may not be involved in all intrastate conflicts. In many countries, people give preference to members of their own political party, ideology, tribe, religion, and ethnicity. Even when these groups are not fighting against the state, there are many intergroup conflicts where two or more groups fight with each other and such conflict is recorded as non-sate conflicts when the threshold of twenty-five battle-related deaths is crossed. Consequently, we use the UCDP non-state conflict dataset[4] , Version 2.5-2014, 1989-2013 (Sundberg et al. 2012), to obtain a measure of non-state conflict incidence.

2.2.2 Main explanatory variable (variation in potential average crop yield)

As discussed earlier, our independent variable in both, reduced-form cross-country, and repeated cross-country analyses is variation in potential crop yield. This index captures variation in potential

[4]For example, between paramilitary United Self-Defense Forces of Colombia (AUC) and the guerrilla Revolutionary Armed Forces of Colombia (FARC) in Colombia, or between the competing rebel groups National Patriotic Front of Liberia (NPFL) and United Liberation Movement of Liberia for Democracy (ULIMO) in Liberia ; clashes between supporters of the All Nigeria People's Party and supporters of the People's Democratic Party in Nigeria; between groups that define themselves along identity lines, be it ethnic, clan, religious, national, or tribal identities. Fighting between these types of groups captures what is commonly referred to as communal conflict, such as Hindu-Muslim riots in India.

crop yield (measured in millions of kilocalories per hectare per year) across different regions in each country. It addresses the limitations of other available weight-based agricultural yield indices. For example, the land productivity index constructed by Ramankutty et al. (2002), does not reflect the fact that equally suitable land can have a large variation in crop yield since caloric-intensive crops may not be cultivated in some regions (see Galor and Özak, 2015).

The data on crop yield are taken from Galor and Özak (2016), who construct their dataset from the Global Agro-Ecological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). GAEZ supplies global yield and growth cycle estimates for 48 crops in grids with cells size of 5'x5' (approximately 100 km2). GAEZ supplies crop yield estimates for each crop based on three alternative levels of input – high, medium and low – and two feasible sources of water supply – rain-fed water supply and irrigation. The FAO dataset supplies potential yield estimates for each crop in each agro-climatic grid while accounting for the effect of temperature and moisture on the growth of the crop.

The crop yield index is constructed using low level of inputs and rain-fed agriculture cultivation methods in order to account for the agricultural conditions that were prevalent during the pre-industrial era. This reflects farming practices prevalent during the early stages of development. Furthermore, the crop yield data are based on agroclimatic conditions that are independent of any human intervention. This, therefore, mitigates any potential endogeneity pertaining to agricultural inputs, irrigation method, and soil quality.

Each crop yield in the GAEZ data (measured in tons per hectare per year) is converted by Galor and Özak (2016) into caloric yield (millions of kilocalories per hectare per year) to represent the nutritional variation across crops and to compare crop yields among different crops. The US Department of Agriculture Nutrient Database for Standard Reference provides data on the caloric content of various crops. Using these measures, the estimated average regional crop yield reflects the average regional levels of two variables – crop yield and crop cycle – among crops that indicate the average caloric yield in each cell. Our analysis focuses on the averages across cells where the average potential crop yield is positive. We take logs on the standard deviation of potential (average) crop yield for all the grids in a country and use it as our measure of variation in potential (average) crop yield.

Note that the survival of human society is not feasible in regions with no crop yield and hence, the consideration of cells with positive average potential crop yield is more reasonable. Nevertheless, the correlation between our preferred measure and the alternative measure that includes cells with zero crop yield is 0.93 in our sample, suggesting that excluding cells with no crop yield is unlikely to drive the results significantly. It is also worth highlighting that potential crop yield is positively correlated with actual crop yield at the cell level. However, actual crop is likely to be affected by human intervention and the event of the conflict itself. Hence, using actual crop yield data may give rise to some endogeneity problems. Moreover, GAEZ provides actual crop yield data only for the year 2000, which in our view, are not sufficiently exogenous.

Figure 1 plots the distribution of variation in crop yield for all the countries in the sample. The bottom decile of countries has a variation below 62.60, while the top decile has a variation of over 674.56. Countries in Middle-East Asia and North Africa (e.g., Israel, Qatar, Saudi Arabia, UAE, and Oman) typically have very low levels of variation in crop yield whereas some mountainous or geographically large countries such as the U.S., Argentina, Nepal, Bhutan, and China, are at the other end of the spectrum. Figure 2 shows that conflict incidence and variation in crop yield displays a positive relationship.

[Figure 1]

[Figure 2]

2.2.3 Control variables

Mean potential (average) crop yield is included in all regressions. The computation of this variable is similar to the way variation in potential (average) crop yield is constructed, except for the fact that it represents the mean of potential (average) crop yield over all the cells in a country. Several other control variables which may potentially confound the relationship between variation in crop yield and conflict incidence are included in the regressions. This includes terrain ruggedness, absolute latitude, elevation, distance to nearest waterway, precipitation, landlockedness, mean elevation, temperature and the percentage of population living in tropical zones. We also include continent dummies to account for any omitted variable bias of time-invariant continent-specific geographical, cultural and historical characteristics. Summary statistics and the correlations of the variables are given in Table 1.

[Table 1]

3 Empirical Results

3.1 Cross-country analysis of intrastate conflict

Table 2 presents our findings for the influence of variation in crop yield on conflict incidence in a cross-section of 135 countries. The analysis here makes use of data on the ongoing conflict incidence dimension of civil conflict over the period from 1960 to 2014.

[Table 2]

Column (1) uses no control variable and shows that the unconditional influence of variation in crop yield on conflict incidence is positive and significant at 1% level. Measured in a standardized form, a one standard deviation change in (logs) variation in crop yield is expected to result in 0.17 standard deviations increase in conflict incidence. We introduce mean crop yield in column (2). The results show that the influence of variation in crop yield on conflict incidence, after accounting for the influence of mean crop yield, remains highly statistically significant. We include standard geographic controls in column (3) and the coefficient of crop yield variation remains statistically significant. Finally, although we have accounted for various country-level geographical features, there is still possibility that the results may be biased due to the failure to control for other unobserved characteristics. Hence, we include continent fixed effects in column (4) and the results remain robust. In all cases, we do not find any significant effect of average crop yield. The partial regression plot in Figure 3 shows the relationship between conflict incidence and variation in crop yield. As expected, the partial regression line shows a positive and statistically significant influence of variation in crop yield on conflict incidence, after controlling for other geographical controls and continent fixed effect in column (4) of Table 2.

[Figure 3]

3.2 Robustness to alternative definitions and samples of conflict

An implicit assumption of our analysis is that the natural endowments of a region have not changed much over time and hence our results can explain the occurrence of conflicts in the past over a long duration and, the current level of conflict should be correlated with the historic level of conflict. To check the validity of this assumption, we use the total number of historic internal conflicts compiled by Brecke (1999). Historic conflicts since 1500AD appear to be significantly correlated with our outcome variable - the recent conflicts since 1960. When we use the total number of conflict since 1500AD as the outcome variable in column (1) of Table 3, the results indicate that the influence of variation in crop yield on conflict over the last five centuries remains consistent.

[Table 3]

We have used the number of conflict incidence with at least 25 battle-related deaths within a year as our outcome variable so far. There could be some concerns regarding the way conflict intensity is defined (i.e., the number of 5-year periods that has at least one conflict over the period 1960-2014). For instance, a country that has experienced only one incident of conflict in a particular five-year period will be regarded as having the same extent of conflict intensity as another country having experienced conflicts every year over the same period.

We address this problem in subsequent robustness tests by using various alternative definitions and data sources to measure conflict intensity. Accordingly, to ensure that our results are not driven by the way the outcome variable is defined, we use the fraction of years that had incidence of conflict over the period 1960-2014 to mitigate the concern that the intensity of conflicts is smoothed out over a five-year period. Our results shown in column (2) are not affected by this consideration.

Further, there are other ways to count the existence of a conflict in a country based on alternative thresholds for the number of deaths in a conflict. We use *confinc_mid* and *confinc_war* as alternative measures of conflict incidence, where the former captures medium conflict (less than 1000 battle-related deaths in a given year but at least cumulative 1000 battle-related deaths since the beginning of conflict) and the latter captures major conflict (at least 1000 battle-related deaths in a given year) respectively, as defined earlier.

The estimation results for using *confinc_mid* and *confinc_war* as outcome variables are presented in columns (3) and (4), respectively. The results indicate that the influence of variation in crop yield on medium conflict incidence (column 3) and major conflict incidence (column 4) is statistically highly significant in both cases. Notice that the value of the adjusted R2 in column (4) drops substantially. One possible explanation is that the occurrence of major conflicts is likely to be largely determined by idiosyncratic shocks not captured by our model. For this reason, we believe that *confinc*, the measure of conflict incidence that includes low-intensity conflict with at least 25 but less than 1000 battle-related deaths in a year, is a more appropriate outcome measure for the purpose of our study.

Next, we perform additional robustness checks by using two measures on the onset of new conflicts as our alternative outcome variables. These measures, namely *onset2* and *onset5*, suffer from the problem that they cannot fully distinguish between new episodes of conflict and ongoing conflict when battle-related deaths exceed the threshold of 25 in a given year. Bearing this limitation in mind, we prefer our baseline approach to focus on the frequency of new conflict. Columns (5) and (6) present the results of estimating the influence of variation in crop yield on the frequency of the onset of new conflict using *onset2* and *onset5* as the dependent variable, respectively. In both cases, the effect of variation in crop yield is still statistically significant at 1% level.

In column (7), we present the results for regressing the onset of new ethnic conflict on the variation of crop yield. It is evident that the results are qualitatively very similar to what we obtained earlier. Lastly, the influence of variation in crop yield on the incidence of non-state conflicts where none of the sides involved in the conflict is a state, is tested in column (8). The results again remain largely intact to this alternative outcome measure.

3.3 Robustness to the method of estimation

Our baseline outcome variable measures the intensity of conflict incidence in a country on an 11-point scale. Moreover, given that our dependent variable takes positive integer values with a large number of zero values, the data are over dispersed (the variance of 10.32 is much larger than the mean value of 2.54). Consequently, we also perform count data analysis, i.e., Poisson and Negative binomial regressions, as suggested by Cameron and Trivedi (1986) and Wooldridge (2010), to provide a robustness check of our results. Columns (9) and (10) of Table 3 report the findings using these alternative estimators, respectively, and indicate that our earlier results prevail.

3.4 Robustness to controlling for the effects of early development

Our results may also be biased due to the failure to control for some influential early development effects emphasized in the literature. In particular, Ahlerup and Olsson (2012) show that the duration of human settlement since prehistoric times affects the level of ethnolinguistic diversity of the current population. Similarly, Ashraf and Galor (2013) show that the migratory distance from East Africa explains current income levels. Since ethnolinguistic fractionalization and economic development are closely related to conflict, we control for these confounding factors in columns (1) and (2), respectively. Our results reported in Table 4 clearly indicate that variation in crop yield remains highly significant.

[Table 4]

Depetris-Chauvin (2015) shows that areas which have longer state history are less likely to experience conflicts induced by agricultural shocks since the local governments have greater bureaucratic capability. The timing of agricultural transition and biogeography are other dimensions of early development which are closely correlated to agricultural productivity (see Olsson and Hibbs, 2005 and Putterman, 2008 for a discussion). We account for these effects in columns (3) to (5). Lastly, we also control for the effect of early development using a more generic measure – population density in 1500 AD (logged) – in column (6). The last column reports the findings where all these early development indicators are included together. In all cases, coefficients of these indicators are found to be insignificant whereas the coefficients of variation in crop yield remain very precisely estimated.

3.5 Robustness to the inclusion of additional control variables

In this sub-section, we conduct some further robustness checks on the results by considering some additional control variables and report these results in Table 5. First, we augment the baseline specification by adding executive constraints, democracy, autocracy, oil exports as a fraction of total export, and real GDP per capita. As highlighted by Gurr (2000) and Fearon and Laitin (2003), political systems such as democracy and autocracy are strongly related to conflict. Real GDP per capita acts as a proxy for the opportunity cost of engaging in intrastate conflict whereas oil exports as a fraction of total exports are included to capture the effect of natural resources.

[Table 5]

Moreover, we also control for the effects of colonial legacy, following the results of Posner (2003). Controlling for the influence of colonial legacies is important since they may play an important role in stabilizing political institutions in newly independent states, especially given the fact that colonizers arbitrarily defined ethnic groups for administrative purposes and sometimes followed ethnically biased policies. Accordingly, we use colonial history dummies for British, French, Spanish, Portuguese and other major European colonizers. However, there may be some residual influence of legal systems originating from European colonizers, with large differences between those originating from English common law and those originating from Roman civil law. Hence, we also use dummy variables for British Common Law, French civil law, Socialist law, German civil law, and Scandinavian law to capture the major categories of legal origins, as suggested by La Porta et al. (1998). As shown in column (1), including these variables does not change our results in any significant fashion. Essentially, the effect of variation of crop yield remains highly significant, suggesting that our results are not driven by these possible omitted influences.

While we control for the effect of colonial origins in column (1), Olsson (2009) emphasizes the importance of the temporal dimension of colonial rule on institutions due to the fact that colonizers needed time to establish institutions. Controlling for this effect in column (2) does not amend our findings.

Next, we control for several diversity measures which have been shown to influence some development outcomes in the existing literature. These variables are the ethnic fractionalization index of Fearon (2003) (column (3)), the religion fractionalization index of Alesina et al. (2003) (column (4)), and the fractionalization, polarization and Greenberg Gini indices of Esteban et al. (2012) (column (5)). Our results are largely unaffected.

Furthermore, we control for some country-specific geographical characteristics. Fearon and Laitin (2003) find that the mountainous terrain and non-contiguous region at a distance from the center of political power provides a safe haven for rebel groups and facilitates the emergence and perpetuation of civil war. Finally, Letendre et al. (2010) find that the "disease richness" of a country can provide potential ground for higher levels of conflict. The higher "disease richness" (the total number of different types of infectious diseases in a country) of a country can increase its mortality rate and therefore reduce the opportunity cost for individuals to indulge in a civil war (Cervellati et al., 2011). Controlling for these effects in columns (6) and (7), respectively, does not alter our main findings.

Finally, a recent study by Michalopoulos (2012) highlights that diversity in a society may originate from some geographical features of the region. He shows that geographical features, such as variation in elevation and soil quality, can explain ethnolinguistic diversity. Larger countries may tend to have higher variation in crop yield and hence we also control for this effect. We control for these additional geographic features in column (8) but do not find them to be correlated with conflict.

3.6 Robustness to elimination of regions from estimation sample

We also investigate whether our results – specifically, with respect to the influence of variability in crop yield on conflict incidence – are driven by the inclusion of a particular region in our sample. We drop the observations belonging to a particular continent sequentially and re-estimate our baseline model. The results after dropping one continent at a time are presented in Table 6. The estimates indicate that the influence of variability in crop yield is found to be statistically significant at least at 5% level in all cases. Hence, we conclude that our results are not sensitive to the exclusion of observations from any particular continent.

[Table 6]

4 Further Analyses Using Repeated Cross-country Data

In this section, we exploit the time dimension of conflict incidence using quinquennial repeated cross-country data. A similar regression framework is used except that our outcome variable is now a binary variable measuring the potential for an active conflict episode in country i during any given 5-year period. We also account for intertemporal spillover effect of conflict incidence on contemporary conflict incidence by including a lagged dependent variable in our regression model.

Table 7 gives the regression results using a logit estimator. Column (1) reports results for the unconditional influence of variation in crop yield on conflict incidence. The effect is found to be statistically significant at 1% level. We account for time-invariant geographical control variables in column (2) and further control for continent effects in column (3). Column (4) assesses the influence of variation in crop yield on conflict incidence, conditional on time-invariant geographic effects and time-varying institutional and output variables. In all cases, the effect of crop yield variation remains very precisely estimated at 1% level.

[Table 7]

The various outcome variables that we have employed so far reflect the extensive margin effect of variation in crop yield on intrastate conflict events as they use binary measure depending on a threshold number of the battle-related deaths in a year. Hence, we also employ both ordinal and continuous measures of conflict incidence that capture the severity of intrastate conflict to establish the influence of variation in crop yield on the intensive margin of conflict.

The first measure of conflict of intensity that we employ is a non-binary measure of conflict intensity, *confinc_int*, which is based on UCDP arms conflict dataset. Following Esteban et al. (2012), we categorize *confinc_int* into three distinct intensity level – 0, 1, and 2 in any year. When a country is at "Peace", *confinc_int* is assigned a zero value, conflict incidence satisfying *confinc* but not *confinc_war* is assigned a value of 1 whereas conflict incidence satisfying *confinc_war* criteria are assigned a value of 2.

Our second measure of conflict intensity is based on the Major Episodes of Political Violence (MEPV), 1946-2014 dataset (Marshall, 2014). Major episodes of conflicts, according to this database, involve "at least 500 directly-related fatalities and reach a level of intensity in which political violence is both systematic and sustained (a base rate of 100 directly-related deaths per annum)". For each episode of conflict, the MEPV dataset provides the "magnitude score" of the destructive impact, or magnitude of the violent episode of the directly-affected areas on a scale of 1 (smallest) to 10 (greatest). We employ data on the total intrastate conflict that includes civil violence, civil conflict, ethnic violence and ethnic war. These data are reported at the country-year level and to convert them into quinquennial intervals, we create our outcome variable by using the maximum level of conflict intensity from any year in each of the 5-year periods.

Our third measure of the intensity of conflict is a continuous index of social conflict (ISC) computed by Cross-National Time-Series (CNTS) Data Archive (Banks, 2011). The data are taken from Esteban et al., 2012. The period of this conflict severity is limited to between 1960 and 2008, or only ten sub-periods of the 5-year interval (we used eleven sub-periods of 5-year interval earlier). The results in Table 7 show that the influence of variation in crop yield on conflict is still statistically significant even when these alternative measures of conflict are used.

[Table 8]

5 Analysis Using Grid-level Data

In this section, we provide further evidence using data at the grid level. The analysis is performed using a global grid of 2.5 by 2.5 decimal degrees covering from -180 to 180 degrees longitude and from -65 to 85 degrees latitude. To create our outcome variable, we use the latest version (version 5) of UCDP Georeferenced Event Dataset (Sundberg and Melander, 2013). This dataset provides global conflict events (excluding Syria) over the period 1989-2015. A conflict event is defined as: *"An incident where armed force was by an organised actor against another organized actor, or against civilians, resulting in at least 1 direct death at a specific location and a specific date"*. The reporting unit is an individual event and the dataset contains 128,264 events. Figure 4 shows the distribution of conflict across grid cells.

[Figure 4]

It should be highlighted that this dataset is compatible with the one used earlier for cross-sectional analysis. Only conflict events involving at least 25 casualties are reported. The best spatial resolution for the conflict events of this dataset is individual village or town. Three type of events – state based, non-state, and one sided – are reported in the dataset.

As in previous sections, we focus on state-based events where one party in the conflict is the state. This reduces the number of conflict events in the dataset to 89,943. We match these violent events, using latitude and longitude of the events, with our grid of 2.5 by 2.5 decimal to identify corresponding grid cells for each event over the period 1989-2015. Two outcome variables are then created. The first variable indicates whether there was an incidence of conflict in a given grid cell (labeled as *"presence of conflict incidence"*). The second is constructed using an approach similar to that of *confinc*. We divide the whole period 1989-2015 into six 5-year sub-periods (the last sub-period covers only 2 years). We indicate conflict intensity as 1 if there is at least one event of conflict in that grid cell in the 5-year window and 0 otherwise. The aggregate number is our second measure and we label it as *"frequency of conflict incidence"*. The grid-level geographical data are obtained from Michalopoulos (2012).

[Table 9]

The results for using grid-level data are reported in Table 9. As is evident, variation in crop yield is found to have a significant effect on conflict, regardless of which outcome measure is used and whether control variables and continent fixed effects are included.

6 Summary and Conclusions

Agroecology in general, and geography and climate in particular, has become an important topic of research in economics as part of a broader agenda to understand the origins of both, intrastate and interstate conflict. Nevertheless, there is ambiguity in identifying the specific aspect of geography and climate that have significant influence over the likelihood of conflict incidence. Recent studies indicate that resources available in a region and the opportunity cost of the people to participate in violence play a key role in the onset and persistence of conflict incidence.

This study contributes to this growing literature by identifying the role of crop yield (measured in millions of kilocalories per hectare per year) variation as an important source of conflict. Most of the conflicts were initiated to achieve dominance of power over people from another ethnicity, religion, race or gender in order to control lands with high crop yield potential. There can be various plausible mediating channels to explain this relationship. As different groups settle in different parts of a country, their economic prosperity should be positively influenced by the agricultural productivity in that region. Consequently, a country with higher variation in crop yield is expected to have higher ethnic inequality, and this in turn induces more conflicts. Similarly, the formation of cultures of aggression and violence is also likely to occur when societies depend on agriculture for survival. Although, many societies are not as dependent on farming today as they were a few centuries ago the culture of violent behavior developed centuries ago may transmit across generations. As this violent behavior continues to persist, one should expect to observe more intrastate conflict in a country that has higher variation in crop yield across regions.

Our empirical findings provide support for this hypothesis. Specifically, we find a variation in crop yield to have a significant effect on conflict incidence. These findings are robust to the use of alternative outcome variables, measures of conflict, the inclusion of a number of possible confounders, and alternative empirical regression methods. The results are also consistent when grid cell level data are used.

References

Ahlerup, P. & Olsson, O. (2012). 'The roots of ethnic diversity', Journal of Economic Growth 17, 71-102.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. & Wacziarg, R. (2003). 'Fractionalization', Journal of Economic Growth 8, 155-194.

Alesina, A., Giuliano, P. & Nunn, N. (2013). 'On the Origins of Gender Roles: Women and the Plough', Quarterly Journal of Economics 128, 469-530.

Alesina, A., Michalopoulos, S. & Papaioannou, E. (2016). 'Ethnic Inequality', Journal of Political Economy 124, 428-488.

Anderson, C.A. (1987). 'Temperature and aggression: effects on quarterly, yearly, and city rates of violent and nonviolent crime', Journal of personality and social psychology 52, 1161.

Ang, J.B. (2015). 'What Drives the Historical Formation and Persistent Development of Territorial States?' Scandinavian Journal of Economics 117, 1134–1175.

Angrist, J.D. & Kugler, A.D. (2008). 'Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia', Review of Economics and Statistics 90, 191-215.

Arbatli, C.E., Ashraf, Q.H. & Galor, O. (2015). 'The Nature of Conflict', NBER Working Papers 21079.

Ashraf, Q. & Galor, O. (2013). 'The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development', American Economic Review 103, 1-46.

Ashraf, Q. & Michalopoulos, S. (2015). 'Climatic Fluctuations and the Diffusion of Agriculture', Review of Economics and Statistics 97, 589-609.

Blattman, C. & Miguel, E. (2010). 'Civil War', Journal of Economic Literature 48, 3-57.

Bockstette, V., Chanda, A. & Putterman, L. (2002). 'States and Markets: The Advantage of an Early Start', Journal of Economic Growth 7, 347-369.

Brückner, M. (2010). 'Population Size and Civil Conflict Risk: Is there a Causal Link?' The Economic Journal 120, 535-550.

Cameron, A.C. & Trivedi, P.K. (1986). 'Econometric Models Based on Count Data. Comparisons and Applications of Some Estimators and Tests', Journal of Applied Econometrics 1, 29-53.

Cederman, L.-E., Weidmann, N.B. & Gleditsch, K.S. (2011). 'Horizontal Inequalities and Ethnonationalist Civil War: A global Comparison', American Political Science Review 105, 478-495.

Cervellati, M., Sunde, U. & Valmori, S. (2011). 'Disease Environment and Civil Conflicts', IZA Working Paper 5614.

Chaney, E. (2013). 'Revolt on the Nile: Economic Shocks, Religion, and Political Power', Econometrica 81, 2033-2053.

Collier, P. & Hoeffler, A. (1998). 'On Economic Causes of Civil War', Oxford Economic Papers 50, 563-573.

Collier, P. & Hoeffler, A. (2004). 'Greed and Grievance in Civil War', Oxford Economic Papers 56, 563-595.

Collier, P. & Sambanis, N. (2002). 'Understanding Civil War: a New Agenda', Journal of

Conflict Resolution 46, 3-12.

Deininger, K. (2003). 'Causes and Consequences of Civil Strife: Micro-level Evidence from Uganda', Oxford Economic Papers 55, 579-606.

Dell, M., Jones, B.F. & Olken, B.A. (2012). 'Temperature Shocks and Economic Growth: Evidence from the Last Half Century', American Economic Journal: Macroeconomics 4, 66-95.

Depetris-Chauvin, E. (2015). 'State history and contemporary conflict: Evidence from Sub-Saharan Africa', Available at SSRN: https://ssrn.com/abstract=2679594.

Depetris-Chauvin, E. & Özak, Ö. (2015). 'Population Diversity, Division of Labor and the Emergence of Trade and State', Available at SSRN: https://ssrn.com/abstract=2679382.

Depetris-Chauvin, E. & Özak, Ö. (2016). 'The Origins and Long-Run Consequences of the Division of Labor', Available at SSRN: https://ssrn.com/abstract=2827328.

Esteban, J., Mayoral, L. & Ray, D. (2012). 'Ethnicity and Conflict: An Empirical Study', American Economic Review 102, 1310-1342. Fearon, J.D. (2003). 'Ethnic and Cultural Diversity by Country', Journal of Economic Growth 8, 195-222.

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

Galor, O. & Özak, Ö. (2015). 'Land Productivity and Economic Development: Caloric Suitability vs. Agricultural Suitability', Brown University, Department of Economics Working Papers 2015-5.

Galor, O. & Özak, Ö. (2016). 'The Agricultural Origins of Time Preference', American Economic Review 106, 3064-3103.

Gurr, R.T. (1970) Why Men Rebel: Princeton, NJ: Princeton University Press).

Gurr, T.R. (2000). 'Ethnic Warfare on the Wane', Foreign Affairs 79, 52-64.

Hibbs, D.A., Jr. & Olsson, O. (2004). 'Geography, Biogeography and Why Some Countries Are Rich and Others Poor', Proceedings of the National Academy of Sciences of the United States 101, 3715-3720.

Hsiang, S.M., Meng, K.C. & Cane, M.A. (2011). 'Civil Conflicts are Associated with the Global Climate', Nature 476, 438-441.

Jimenez-Ayora, P. & Ulubaşoğlu, M.A. (2015). 'What underlies weak states? The role of terrain ruggedness', European Journal of Political Economy 39, 167-183.

La Porta, R., Silanes, Florencio L.d., Andrei Shleifer & Robert W. Vishny (1998). 'Law and Finance', Journal of Political Economy 106, 1113-1155.

Letendre, K., Fincher, C.L. & Thornhill, R. (2010). 'Does Infectious Disease Cause Global Variation in the Frequency of Intrastate Armed Conflict and Civil War?' Biological Reviews 85, 669-683.

McEvedy, C. & Jones, R. (1978) Atlas of World Population History (New York: Facts on File).

Michalopoulos, S. (2012). 'The Origins of Ethnolinguistic Diversity', American Economic Review 102, 1508-1539.

Miguel, E., Satyanath, S. & Sergenti, E. (2004). 'Economic Shocks and Civil Conflict: An

Instrumental Variables Approach', Journal of Political Economy 112, 725-753.

Murray, D.R. & Schaller, M. (2010). 'Historical Prevalence of Infectious Diseases within 230 Geopolitical Regions: A Tool for Investigating Origins of Culture', Journal of Cross-Cultural Psychology 41, 99-108.

Nunn, N. & Puga, D. (2012). 'Ruggedness: The Blessing of Bad Geography in Africa', Review of Economics and Statistics 94, 20-36.

Nunn, N. & Qian, N. (2014). 'US Food Aid and Civil Conflict', American Economic Review 104, 1630-1666.

Olsson, O. (2009). 'On the Democratic Legacy of Colonialism', Journal of Comparative Economics 37, 534-551.

Olsson, O. & Hibbs, D.A. (2005). 'Biogeography and Long-Run Economic Development', European Economic Review 49, 909-938.

Olsson, O. & Paik, C. (2016). 'Long-Run Cultural Persistence: Evidence from the Neolithic Revolution', Journal of Development Economics 122, 197-213.

Pettersson, T. & Wallensteen, P. (2015). 'Armed Conflicts, 1946–2014', Journal of Peace Research 52, 536-550. Posner, D.N. (2003). 'The Colonial Origins of Ethnic Cleavages: The Case of Linguistic Divisions in Zambia', Comparative Politics 35, 127-146.

Putterman, L. (2006). 'Agricultural Transition Year Country Data Set', Brown University.

Putterman, L. (2008). 'Agriculture, Diffusion and Development: Ripple Effects of the Neolithic Revolution', Economica 75, 729-748.

Putterman, L. (2012). 'State Antiquity Index (Statehist) Version 3.1 (http://www.econ.brown.edu/fac/louis_putterman/)'.

Ramankutty, N., Foley, J.A., Norman, J. & McSweeney, K. (2002). 'The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change', Global Ecology and Biogeography 11, 377-392.

Schlenker, W. & Lobell, D.B. (2010). 'Robust Negative Impacts of Climate Change on African Agriculture', Environmental Research Letters 5, 014010.

Sundberg, R., Eck, K. & Kreutz, J. (2012). 'Introducing the UCDP Non-State Conflict Dataset', Journal of Peace Research 49, 351-362.

Sundberg, R. & Melander, E. (2013). 'Introducing the UCDP Georeferenced Event Dataset', Journal of Peace Research 50, 523-532.

Van de Vliert, E., Schwartz, S.H., Huismans, S.E., Hofstede, G. & Daan, S. (1999). 'Temperature, Cultural Masculinity, and Domestic Political Violence A Cross-National Study', Journal of Cross-Cultural Psychology 30, 291-314.

Wimmer, A., Cederman, L.-E. & Min, B. (2009). 'Ethnic Politics and Armed Conflict: a Configurational Analysis of a New Global Dataset', American Sociological Review 74, 316-337.

Wooldridge, J.M. (2010) Econometric Analysis of Cross Section and Panel Data (Cambridge, MA: MIT Press).

Zhang, D.D., Brecke, P., Lee, H.F., He, Y.-Q. & Zhang, J. (2007). 'Global Climate Change, Sciences 104, 19214-19219.

Table 1: Summary statistics and correlations

Summary Statistics	No. of observations	Mean	Standard deviation	Minimum	Maximum
Conflict incidence	135	2.3	2.9	0	11
Variation in crop yield	135	369.6	252.3	9.2	1127.3
Mean crop yield	135	1207.4	649.7	0.5	2742.8
Terrain ruggedness	135	0.2	0.2	0.0	1.0
Absolute latitude	135	0.3	0.2	0.0	0.7
Elevation	135	0.6	0.5	0.0	2.7
Distance to waterways	135	0.3	0.5	0.0	2.4
Precipitation	135	86.6	59.3	2.9	260.0
Landlockedness	135	0.3	0.4	0.0	1.0
Temperature	135	17.9	8.5	-7.9	28.6
% of pop in tropical zones	135	0.3	0.4	0.0	1.0

Correlation Matrix	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Conflict incidence	1										
(2) Variation in crop yield	0.14	1									
	(0.09)										
(3) Mean crop yield	-0.13	-0.05	1								
	(0.12)	(0.53)									
(4) Terrain ruggedness	0.21	0.43	0.06	1							
	(0.81)	(0.00)	(0.47)								
(5) Absolute latitude	-0.32	0.21	-0.02	0.11	1						
	(0.00)	(0.02)	(0.84)	(.022)							
(6) Elevation	0.07	0.58	-0.1	0.68	-0.02	1					
	(0.43)	(0.00)	(0.25)	(0.00)	(0.79)						
(7) Distance to waterways	0.09	0.20	-0.37	-0.04	0.06	0.37	1				
	(0.29)	(0.02)	(0.00)	(0.64)	(0.51)	(0.00)					
(8) Precipitation	0.16	-0.21	0.35	0.06	-0.50	-0.14	-0.34	1			
	(0.07)	(0.01)	(0.00)	(0.49)	(0.00)	(0.10)	(0.00)				
(9) Landlockedness	-0.13	0.24	0.1	0.25	0.09	0.46	0.46	-0.22	1		
	(0.15)	(0.00)	(0.24)	(0.00)	(0.28)	(0.00)	(0.00)	(0.01)			
(10) Temperature	0.31	-0.34	-0.02	-0.26	-0.89	-0.25	-0.23	0.35	-0.19	1	
	(0.00)	(0.00)	(0.85)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.03)		
(11) % of pop in tropical zones	0.23	-0.39	0.22	-0.20	-0.74	-0.21	-0.16	0.76	-0.25	0.62	1
	(0.00)	(0.00)	(0.01)	(0.02)	(0.00)	(0.01)	(0.06)	(0.00)	(0.00)	(0.00)	

Notes: Conflict incidence is the number of five year sub-periods, between 1960 and 2014, that has at least one ongoing conflict of any intensity. Variation in crop yield and mean crop yield are the standard deviation and mean, respectively, of potential (average) crop yield (measured in millions of kilo calories per hectare per year) across different regions in each country. Significance level are reported in the brackets.

Table 2: Main regression results (cross-sectional OLS)

	(1)	(2)	(3)	(4) (Baseline)
Dependent variable:	Conflict incidence (*confinc*)			
Variation in crop yield, log	0.17***	0.20**	0.34***	0.47***
	(2.67)	(2.52)	(3.46)	(4.71)
Mean crop yield, log		-0.07	-0.07	-0.08
		(-1.06)	(-0.71)	(-0.86)
Rugged terrain, %			0.04	-0.02
			(0.35)	(-0.13)
Absolute Latitude			0.52*	0.08
			(1.79)	(0.26)
Elevation			0.14	-0.00
			(0.97)	(-0.00)
Distance to waterways			0.29**	0.23*
			(2.60)	(1.69)
Precipitation			0.14	0.07
			(0.84)	(0.44)
Landlockedness			-0.25***	-0.24***
			(-2.78)	(-2.73)
Temperature			0.80***	0.57**
			(3.10)	(2.14)
% pop in tropical zones			0.15	0.16
			(0.94)	(1.04)
Observations	135	135	135	135
Adjusted R-squared	0.02	0.02	0.20	0.30
Continent dummies	No	No	No	Yes

Notes: This table presents standardized beta coefficients of the influence of variation in crop yield on conflict incidence in reduced-form. The continent dummies are Africa, Asia, Europe, North America, South America, with Oceania as the excluded group. An intercept is included in all the regressions but not reported for brevity. *** denotes statistical significance at 1% level, ** at the 5% level and, * at the 10% level. Robust t-statistics are given in parentheses

Table 3: Robustness to alternative methods and measures of conflict

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. Var. = (estimator)	confict_hist (OLS)	confinc_frac (OLS)	confinc (Poisson)	confinc (Neg. Binomial)	confinc_mid (OLS)	confinc_war (OLS)	onset2 (OLS)	onset5 (OLS)	onset_eth (OLS)	confinc_ns (OLS)
Var in crop yield, log	0.24***	0.46***	0.23***	0.24***	0.39***	0.29**	0.41***	0.43***	0.33***	0.46***
	(2.71)	(4.25)	(4.46)	(4.34)	(3.28)	(2.49)	(3.76)	(4.14)	(2.90)	(4.64)
Mean crop yield,log	-0.01	-0.08	-0.05	-0.07	-0.04	0.05	-0.01	-0.06	0.09	-0.04
	(-0.09)	(-0.91)	(-0.95)	(-1.14)	(-0.42)	(0.69)	(-0.15)	(-0.58)	(1.05)	(-0.41)
Observations	135	135	135	135	135	135	135	135	135	131
Pseudo / Adjusted R^2	0.10	0.23	0.28	0.12	0.16	0.09	0.22	0.22	0.19	0.25
Baseline controls	YES	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	YES
Continent FE	YES	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	YES

Notes: Dependent variable in columns (1) and (2) are *conflict_hist,* the total number conflicts during the period 1500-2010, and *confinc_frac,* the fraction of years that had incidence of conflict over period 1960-2014. Columns (3) to (4) present results using Poisson and negative binomial regression methods, respectively. Pseudo R-squared values are reported for these regressions. Columns (5) to (10) present standardized beta coefficients using ordinary least square method. *confinc* is number of five year sub-periods, between 1960 and 2014, that had at least one ongoing conflict of any intensity; *confinc_mid* is number of five year sub-periods, between 1960 and 2014, that had at least one ongoing conflict where battle-related deaths were less than one thousand in a single year but cumulative battle-related deaths exceeded one thousand since the beginning of the episode of conflict in any year within a 5-year sub-period; *confinc_war* is number of five year sub-periods, between 1960 and 2014, that had at least one ongoing conflict where battle-related deaths exceeded one thousand in a single year in any year within the 5-year sub-period; *onset2* and *onset5* are the number of five-year sub-periods, between 1960 and 2014, that had at least one onset of new conflict episode after two and five years of continuous peace respectively; *onset_eth* is the aggregate of new ethnic conflicts between 1960 and 2010; *confinc_ns* are number of five year sub-periods, between 1960 and 2014, that had at least one ongoing non-state conflict respectively; Adjusted R-squared values are reported for these regressions. Absolute latitude, elevation, distance to sea, average precipitation, average temperature, ruggedness, landlockedness, and % population in tropical zones are included as additional controls in all regressions. The continent dummies are Africa, Asia, Europe, North America, South America and Oceania. An intercept is included in all the regression but not reported for brevity of space. ***

Table 4: Robustness to controlling for the effects of early development

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable			Conflict incidence (*confinc*)				
Variation in crop yield, log	0.42***	0.47***	0.39***	0.46***	0.37***	0.46***	0.35***
	(4.34)	(4.61)	(3.75)	(4.45)	(3.89)	(4.55)	(3.14)
Mean crop yield	-0.06	-0.07	-0.07	-0.06	-0.08	-0.10	-0.05
	(-0.62)	(-0.71)	(-0.86)	(-0.63)	(-0.96)	(-0.96)	(-0.43)
Origtime	-0.07						-0.01
	(-0.42)						(-0.04)
Migratory distance from East Africa		-0.41					-0.66
		(-1.04)					(-1.24)
State history up to 1500 AD			0.08				0.10
			(0.58)				(0.57)
Years since agricultural transition				0.09			-0.11
				(0.59)			(-0.49)
Biogeography					0.07		-0.02
					(0.46)		(-0.09)
Population density in 1500 AD (log)						0.09	-0.00
						(0.76)	(-0.00)
Observations	128	130	123	126	121	128	113
Adjusted R-squared	0.31	0.31	0.30	0.30	0.31	0.31	0.32
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports standardized beta coefficients using OLS. All regressions include the baseline controls, including absolute latitude, elevation, distance to sea, average precipitation, average temperature, ruggedness, landlockedness, and % population in tropical zones. The continent dummies are Africa, Asia, Europe, North America, South America and Oceania. An intercept is included in all the regression but not reported for brevity of space. *** denotes statistical significance at 1% level, ** at the 5% level and, * at the 10% level. Robust t-statistics are given in parenthesis.

Table 5: Robustness to the inclusion of additional covariates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable				Conflict incidence (*confinc*)				
Variation in crop yield, log	0.39***	0.33**	0.47***	0.46***	0.40***	0.41***	0.42***	0.48***
	(4.15)	(2.31)	(4.77)	(4.42)	(3.91)	(3.87)	(4.04)	(4.20)
Mean crop yield, log	-0.12	-0.15	-0.09	-0.07	-0.16	-0.08	-0.03	-0.11
	(-1.05)	(-1.07)	(-0.87)	(-0.64)	(-1.54)	(-0.83)	(-0.26)	(-0.80)
Executive control	0.34							
	(1.48)							
Autocracy	-0.11							
	(-0.58)							
Democracy	-0.21							
	(-1.07)							
% Oil export in total export	0.06							
	(0.71)							
GDP per capita, log	-0.31**							
	(-2.52)							
Colonial history (F-statistic)	1.98							
	[p=0.102]							
Legal origin (F-statistic)	1.98							
	[p=0.103]							
Duration of colonial experience		0.19						
		(1.55)						
Ethnic fractionalization (Fearon, 2003)			0.10					
			(1.05)					
Religion fractionalization (Alesina et. al., 2003)				-0.06				
				(-0.67)				
Fractionalization (Estaben, 2012)					0.14			
					(1.26)			
Polarization (Estaben, 2012)					0.24**			
					(2.27)			
Greenberg Gini index (Estaben, 2012)					-0.27***			
					(-3.56)			
Mountainous terrain						0.22		
						(1.33)		
Non-contiguous						0.22**		
						(2.24)		
Disease richness							-0.05	
							(-0.24)	
Variation in elevation								0.01
								(0.07)
Mean land quality (soil-based)								0.03
								(0.29)
Land area (1000 km2)								0.09
								(0.66)
Observations	135	75	135	133	118	135	124	130
Adjusted R-squared	0.34	0.36	0.30	0.29	0.39	0.33	0.31	0.29
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports standardized beta coefficients using OLS. All regressions include the baseline controls, including absolute latitude, elevation, distance to sea, average precipitation, average temperature, ruggedness, landlockedness, and % population in tropical zones. The continent dummies are Africa, Asia, Europe, North America, South America and Oceania. An intercept is included in all the regression but not reported for brevity of space. *** denotes statistical significance at 1% level, ** at the 5% level and, * at the 10% level. Unless stated otherwise, robust t-statistics are given in parentheses. The colonial history dummies are Spanish, British, French, and Portuguese. The legal origin dummies are British Common Law, French civil law, Socialist law, German civil law, and Scandinavian law.

Table 6: Robustness to the elimination of regions

	(1)	(2)	(3)	(4)	(5)
Dependent variable			Conflict incidence (*confinc*)		
Variation in crop yield, log	0.40***	0.28**	0.56***	0.49***	0.49***
	(3.35)	(2.36)	(4.24)	(4.82)	(4.60)
Mean crop yield	0.03	-0.13	-0.12	-0.11	-0.07
	(0.26)	(-1.28)	(-0.83)	(-1.06)	(-0.69)
Continent FE	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes
Omitted Continent	Africa	Asia	Europe	North America	South America
Observations	94	99	104	122	124
AdjustedR^2	0.34	0.22	0.26	0.34	0.32

Notes: This table reports standardized beta coefficients using OLS. All regressions include the baseline controls, including absolute latitude, elevation, distance to sea, average precipitation, average temperature, ruggedness, landlockedness, and % population in tropical zones. The continent dummies are Africa, Asia, Europe, North America, South America and Oceania. An intercept is included in all the regression but not reported for brevity of space. *** denotes statistical significance at 1% level, ** at the 5% level and, * at the 10% level. Robust t-statistics are given in parenthesis.

Table 7: Repeated cross-country results (logistic regressions)

	(1)	(2)	(3)	(4)
Dependent variable		Conflict incidence (*confinc*)		
Variation in crop yield, log	0.68***	1.08***	1.47***	1.46***
	(2.86)	(2.90)	(3.86)	(3.81)
Mean crop yield, log	-0.39**	-0.40	-0.54	-0.64
	(-1.98)	(-1.12)	(-1.43)	(-1.63)
Rugged terrain, %		0.48	0.46	0.65*
		(1.26)	(1.14)	(1.73)
Absolute Latitude		1.17	0.53	0.06
		(1.42)	(0.58)	(0.07)
Elevation		0.42	0.04	-0.99*
		(0.96)	(0.08)	(-1.81)
Distance to waterways		1.27***	1.12***	0.90**
		(3.58)	(2.83)	(1.96)
Precipitation		0.20	0.16	-0.08
		(0.48)	(0.35)	(-0.16)
Landlockedness		-1.01***	-0.94***	-0.70*
		(-3.00)	(-2.71)	(-1.85)
Temperature		2.28***	1.80**	0.51
		(3.18)	(2.14)	(0.59)
% pop in tropical zones		0.53	0.68	0.41
		(0.98)	(1.34)	(0.71)
Executive control				0.54
				(1.03)
Autocracy				-0.72**
				(-2.30)
Democracy				-1.07***
				(-2.68)
Oil				0.21
				(0.90)
GDP per capita, log				-0.75*
				(-1.82)
Conflict incidence, lagged	3.01***	2.71***	2.56***	2.40***
	(14.57)	(13.26)	(12.61)	(11.82)
Continent FE	No	No	Yes	Yes
Colonial history FE	No	No	No	Yes
Legal origin FE	No	No	No	Yes
Observations	1157	1157	1157	1117
pseudo R^2	0.31	0.34	0.36	0.37

Notes: This table reports standardized beta coefficients using a logit estimator. All regressions include the baseline controls, including absolute latitude, elevation, distance to sea, average precipitation, average temperature, ruggedness, landlockedness, and % population in tropical zones. The continent dummies are Africa, Asia, Europe, North America, South America and Oceania. An intercept is included in all the regressions but not reported for brevity of space. *** denotes statistical significance at 1% level, ** at the 5% level and, * at the 10% level. Unless stated otherwise, robust z-statistics are given in parentheses. The colonial history dummies are Spanish, British, French, and Portuguese. The legal origin dummies are British Common Law, French civil law, Socialist law, German civil law, and Scandinavian law.

Table 8: Repeated cross-sectional results for conflict incidence severity

	(1)	(2)	(3)
Dependent variable	*confinc_int*	Major Episodes of Political Violence (MEPV)	index of social conflict (ISC)
Variation in crop yield, log	1.02***	0.24***	0.11***
	(3.92)	(3.13)	(3.22)
Mean crop yield, log	-0.24	-0.06	0.02
	(-0.89)	(-0.66)	(0.64)
Conflict intensity, lagged	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes
Observations	1,451	1,278	1,035
(pseudo) R^2	0.32	0.27	0.39

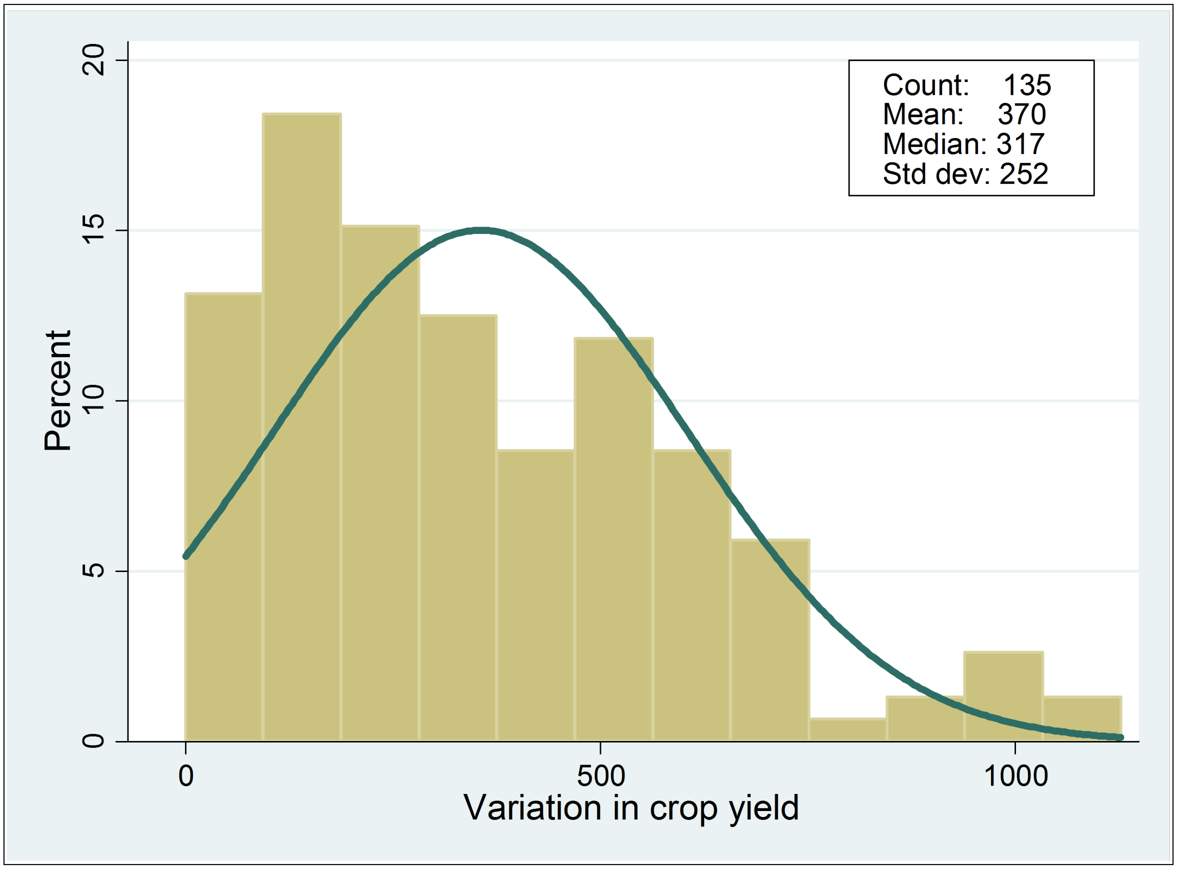
Notes: This table presents standardized coefficients using ordered logit estimation in column (1) and (2), and OLS in column (3). This table reports standardized beta coefficients using OLS. All regressions include the baseline controls, including absolute latitude, elevation, distance to sea, average precipitation, average temperature, ruggedness, landlockedness, and % population in tropical zones. The continent dummies are Africa, Asia, Europe, North America, South America and Oceania. An intercept is included in all the regression but not reported for brevity of space. *** denotes statistical significance at 1% level, ** at the 5% level and, * at the 10% level. Robust z-statistics are given in parentheses.

Table 9: Grid-level analysis

	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
Dependent variable	*Presence of conflict incidence* (Yes = 1)			*Frequency of conflict incidence*		
Variation in crop yield, log	1.17*	0.81*	0.85**	0.21**	0.14**	0.14**
	(1.73)	(1.88)	(1.97)	(2.50)	(2.01)	(2.03)
Mean crop yield, log	-0.33	-0.01	-0.09	0.07	0.17*	0.18*
	(-0.31)	(-0.01)	(-0.18)	(0.66)	(1.80)	(1.88)
Rugged terrain, %		0.02	0.06		0.02	0.02
		(0.13)	(0.35)		(0.47)	(0.48)
Absolute Latitude		2.39**	2.20**		0.36	0.34
		(2.19)	(2.15)		(1.49)	(1.45)
Elevation		1.76***	1.81***		0.44**	0.45**
		(3.25)	(3.67)		(2.26)	(2.38)
Distance to sea		0.15	0.17		-0.06	-0.06
		(0.45)	(0.53)		(-1.17)	(-1.18)
Precipitation		0.38	0.36		0.07	0.07
		(1.08)	(1.00)		(1.28)	(1.21)
Landlockedness		0.07	0.15		0.02	0.03
		(0.46)	(0.92)		(0.90)	(1.00)
Temperature		5.17***	5.33***		1.15*	1.17*
		(2.62)	(2.90)		(1.87)	(1.95)
% pop in tropical zones		0.60	-0.95		-1.20	-1.31
		(0.15)	(-0.22)		(-1.23)	(-1.30)
Baseline controls	No	Yes	Yes	No	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Continent FE	No	No	Yes	No	No	Yes
Observations	1337	1337	1337	1827	1827	1827
Number of clusters	57	57	57	130	130	130
pseudo R^2	0.48	0.50	0.51	-	-	-

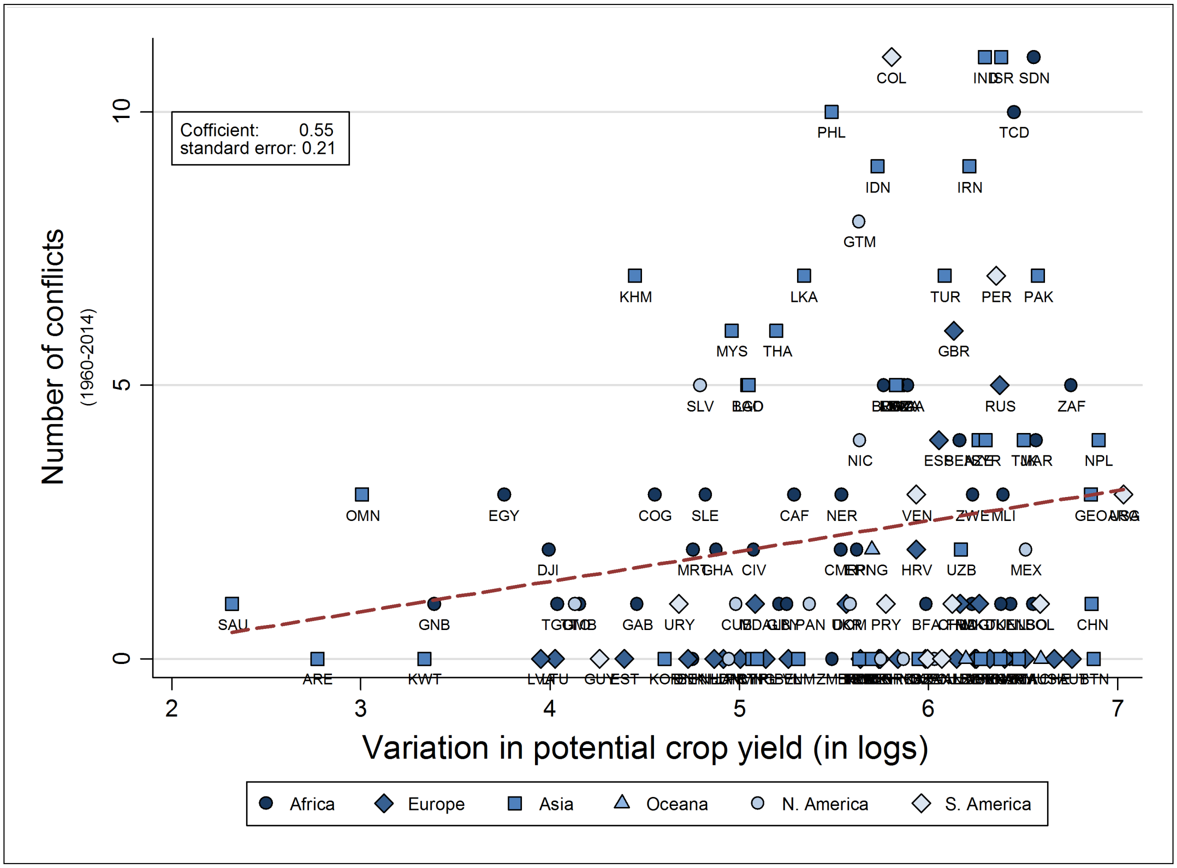
Notes: This table presents standardized beta coefficients of estimation. The dependent variable in all columns (1) is a binary measure indicating whether there has been at least one event of conflict in the grid cell during 1989-2015. The dependent variable in all columns (2) is the number of 5-year periods that has at least one event of conflict event during 1989-2015. We use probit estimation and Poisson estimation methods in columns (1) and (2), respectively. The same set of controls as those used earlier is used for the grid cells. Standard errors are clustered at country level. z-statistics are given in parentheses. * denotes statistical significance at 10% level, ** denotes statistical significance at 5% level, *** denotes statistical significance at 1% level.

Figure 1: Histogram of variation in crop yield



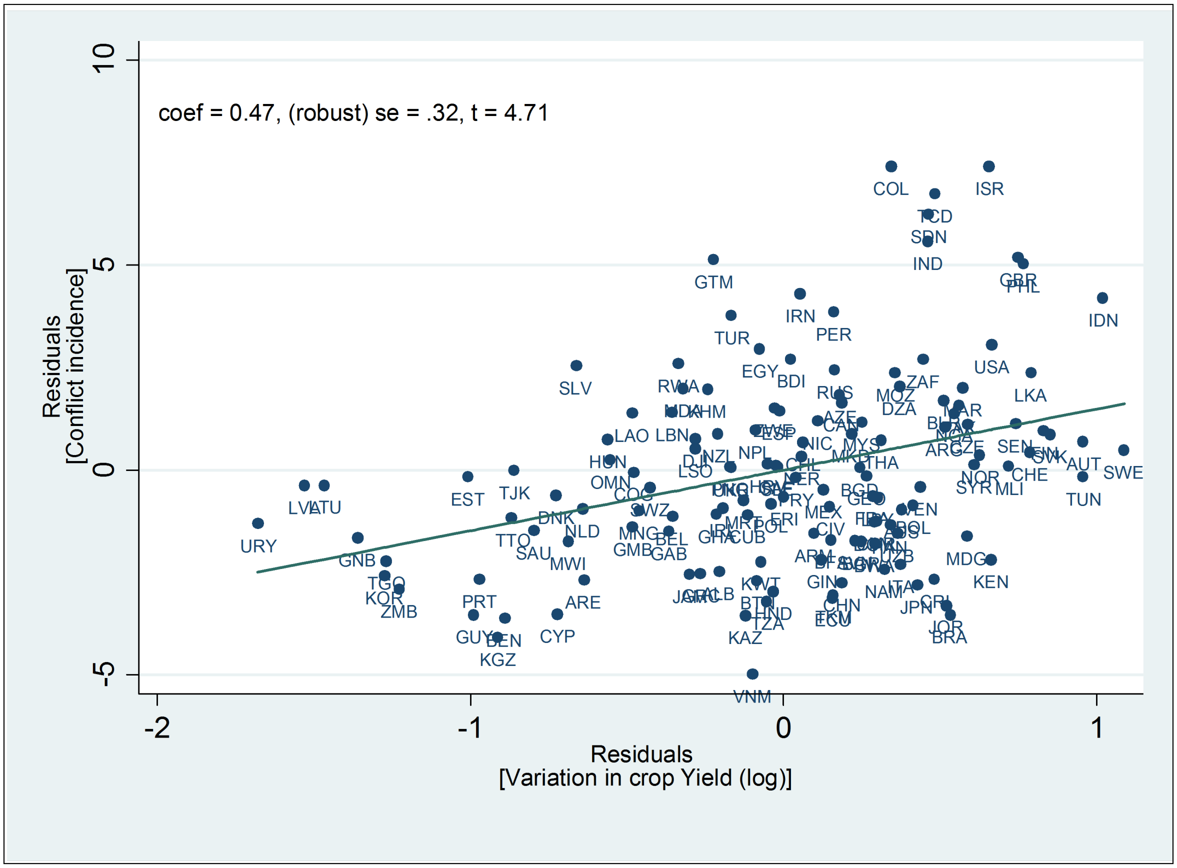
Notes: The horizontal axis measures variation in potential crop yield across the globe, as measured in millions of kilo calories per hectare per year, based on the crops that were available after Colombian exchange (source: Galor and Özak, 2016).

Figure 2: The relationship between conflict incidence and variation in potential crop yield



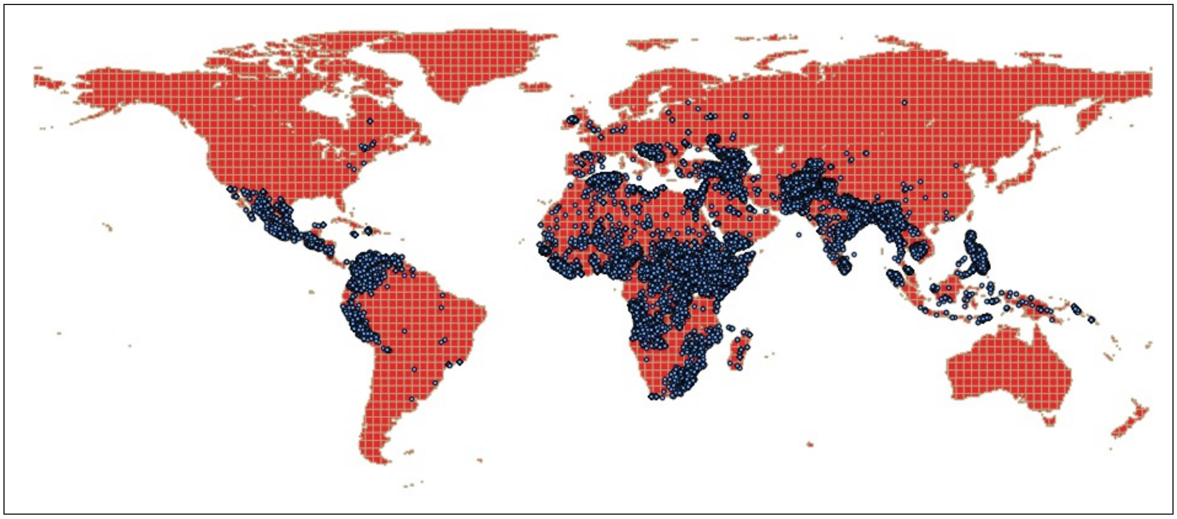
Notes: Conflict incidence is number of five year sub-periods, between 1960 and 2014, that had at least one ongoing conflict of any intensity. The x-axis measures variation in potential crop yield across the globe, as measured in millions of kilo calories per hectare per year, based on the crops that were available after Colombian exchange. The number of countries included is 135 (source: Galor and Özak, 2016).

Figure 3: Relationship between conflict incidence and crop yield variation



Notes: The partial regression line in the diagram illustrates the influence of variation in crop yield on conflict incidence while partialing out the effects of all other control variables, including mean crop yield, rugged terrain, absolute latitude, elevation, distance to waterways, precipitation, landlockedness, temperature, and fraction of population in tropical zones (see column (4) of Table 2). The number of observations is 135.

Figure 4: Conflict data for grid cells of 2.5 by 2.5 decimal degrees



Notes: Longitude extends from -180 to 180 degree, and latitude extends from -65 to 85 degree. Dots represent individual conflict events during the period 1989-2015. The Conflict events data are taken from Sundberg and Melander (2013).

DATA APPENDIX

A. Conflict measures (outcome variables)

Confinc: This variable takes value one when there is at least one ongoing conflict of any intensity, as reported in Armed Conflict Dataset, version 4-2015 (Pettersson and Wallensteen, 2015), in any year within a 5-year sub-period; otherwise it is zero. We use the sum of this variable across 11 sub-periods in cross-section study.

Confinc_mid: This variable takes value one when there is at least one ongoing conflict episode having battle-related deaths of less one thousand in a single year but having cumulative battle-related deaths exceeding one thousand since the beginning of the conflict episode, as reported in Armed Conflict Dataset, version 4-2015 (Pettersson and Wallensteen, 2015), in any year within a 5-year sub-period; otherwise it is zero. We use the sum of this variable across 11 sub-periods in the cross-section study.

Confinc_war: This variable takes value one when there is at least one ongoing conflict episode having battle-related deaths exceeding one thousand in a single year, as reported in Armed Conflict Dataset, version 4-2015 (Pettersson and Wallensteen, 2015), in any year within a 5-year sub-period; otherwise it is zero. We use the sum of this variable across 11 sub-periods in the cross-section study.

Onset2 and onset5: These variables take value one when a new conflict episode starts, as reported in Conflict Onset and Incidence Dataset (Pettersson and Wallensteen, 2015), after two and five years of 'peace', respectively, in any year within a 5-year sub-period; otherwise it is zero. We use the aggregate value of these variables across 11 sub-periods in the cross-section study.

onset_eth: This variable takes value one when a new conflict, that has an ethnic origin, starts in any year; otherwise it is zero. We use the aggregate value of this variable over the period of study. We construct this variable using updated data from Wimmer et al. (2009).

Confinc_ns: This variable is constructed similar to *confinc* defined above, but, unlike the above, no side in the conflict is government or state forces. This variable is constructed using non-state conflict dataset[5] , Version 2.5-2014, 1989-2013 (source: Sundberg et al., 2012).

Confinc_int: This variable indicates the intensity of conflict on a scale of zero to three. When a country is at "Peace", *confinc_int* is assigned a zero value, conflict incidence satisfying *confinc* but not *confinc_war* is assigned a value of 1 whereas conflict incidence satisfying *confinc_war* criteria are assigned a value of 2.

Major Episodes of Political Violence (MEPV): This variable indicates the intensity of conflict on a scale of one (smallest) to ten (highest). We use a maximum of this variable in any year within a 5-year sub-period based on Major Episodes of Political Violence, 1946-2014 dataset (source: Marshall, 2014).

[5] For example, between paramilitary United Self-Defense Forces of Colombia (AUC) and the guerrilla Revolutionary Armed Forces of Colombia (FARC) in Colombia, or between the competing rebel groups National Patriotic Front of Liberia (NPFL) and United Liberation Movement of Liberia for Democracy (ULIMO) in Liberia ; clashes between supporters of the All Nigeria People's Party and supporters of the People's Democratic Party in Nigeria; between groups that define themselves along identity lines, be it ethnic, clan, religious, national, or tribal identities. Fighting between these types of groups captures what is commonly referred to as communal conflict, such as Hindu-Muslim riots in India.

Index of Social Conflict (ISC): This variable indicates the intensity of conflict as computed by Cross-National Time-Series Data Archive for the period from 1960 to 2008 (source: Esteban et al., 2012).

B. Crop yield variation (main explanatory variable)

This index captures variation in potential crop yield (measured in millions of kilocalories per hectare per year) across different regions in a country. Crop yield index is constructed using crop yield (measured in tons per hectare per year) from Global Agro-Ecological Zones (GAEZ) project of Food and Agriculture Organization (FAO) and caloric content of various crops from US Department of Agriculture Nutrient Database for Standard Reference (source: Galor and Özak, 2016).

C. Baseline geographic controls

Ruggedness. The calculation for ruggedness takes a point on the earth's surface and measures the difference in elevation between this point and each of the eight major directions of the compass (north, northeast, east, southeast, south, southwest, west, and northwest). The index at the central point is given by the square root of the sum of the squared differences in elevation between the central point and the eight adjacent points (source: Nunn and Puga, 2012).

Latitude. The absolute value of latitude of a country's centroid, as reported on

http://gothos.info/resources.

Elevation. The average elevation of a country, in thousands of kilometers above sea level, calculated using geospatial data at a 1-degree resolution from the Geographically based Economic data (G-ECON) project.

Distance to waterways. The distance, in thousands of km, from a GIS grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells of a country (source: Harvard University's CID Research Datasets on General Measures of Geography).

Precipitation. The average precipitation of a country, in a unit of 10 millimeters per month over the period from 1961-1990, calculated using geospatial data at a 1-degree resolution from the Goegraphically based Economic data (G-ECON) project.

Landlockedness. This dummy variable indicates whether a country is landlocked or not (source: Global Development Network Growth Database). *Temperature.* The average temperature of a country, in a unit of 10 degree Celsius per month over the period from 1961-1990, calculated using geospatial data at a 1-degree resolution from the Geographically based Economic data (G-ECON) project.

% of population living in tropical zones. The percentage of a country's population in 1995 that resided in areas classified as tropical by the Koppen-Geiger climate classification system (source: Harvard University's CID Research Datasets on General Measures of Geography).

D. Other controls

Origtime. The historical duration of human settlement (in million years) (source: Ahlerup and Olsson, 2012).

Migratory distance from East Africa. The 'Haversine' distance from Addis Ababa to a particular country's modern capital city along a land-restricted route forced through at least one of the five are rescaled so that a higher value indicates greater proximity (source: Ashraf and Galor, 2013).

State history up to 1500 AD. An index of state history covering the period from 1 AD to 1500 AD, scaled to take values between 0 and 1. The dataset was originally introduced by Bockstette et al. (2002), but the current paper uses its latest version, v3.1. (source: Putterman, 2012).

Years since agricultural transition. The number of years elapsed, in 2000 AD, since the transition to agriculture was estimated to occur (in thousand years) (source: Putterman, 2006).

Biogeography. The first principal component of the standardized numbers of domesticable wild plants and animals (source: Hibbs and Olsson, 2004; Olsson and Hibbs, 2005).

Population density in 1500 AD (log). The population in 1500 AD divided by land area (source: McEvedy and Jones, 1978; World Development Indicators, 2012).

Autocracy. An index from the Polity IV project that reflects the degree of institutionalized autocracy. The variable is measured on a scale of 0 to 10, where 10 indicates the largest degree of autocracy (source: Marshall et al., 2012).

Democracy. An index from the Polity IV project that reflects the degree of institutionalized democracy. The variable is measured on a scale of 0 to 10, where 10 indicates the largest degree of democracy (source: Marshall et al., 2015).

Executive constraints. An index from the Polity IV project that reflects the degree of constraints on executives. The variable is measured on a scale of 1 to 7, where 7 indicates the largest amount of constraints (source: Marshall et al., 2015).

Oil. A dummy variable indicating whether the proportion of export revenues from oil is larger than 1/3 (Fearon and Laitin, 2003).

GDP per capita. GDP per capita data at constant price (2005 US$) (source: World Bank Development Indicators 2015).

Colonial history. A classification of a country's colonial origin into British, French, Portuguese, Spanish, or other European (i.e., Dutch, Belgian, and Italian) since 1700. For countries ruled by several colonial powers, the most recent one is taken provided that it was ruled for ten years or longer (source: Nunn and Puga, 2012).

Legal origins. A classification of a country's legal tradition into British common law, French civil law, German civil law, Scandinavian law or Socialist law (source: La Porta et al., 1998).

Duration of colonial history. This is measured as time spent under colonial rule in hundreds of years, i.e. year of independence minus year of colonization, divided by 100 (source: Olsson, 2009).

Ethnic fractionalization. This is defined as the probability that two randomly selected individuals in a country will be from different ethnic groups (source: Fearon, 2003).

Religion fractionalization. This variable is computed as one minus Herfindahl index of religious group shares. It reflects the probability that two randomly selected individuals from population belongs to different religion groups (source: Alesina et al., 2003).

Disease richness. A standardized index of historical pathogen prevalence for nine diseases, including leishmanias, schistosomes, trypanosomes, leprosy, malaria, typhus, filariae, dengue, and tuberculosis (source: Murray and Schaller, 2010).

Fractionalization. This index is constructed as: $F = \Sigma_{i=1}^{m} n_i(1-n_i)$, where n_i is the population share of group i and m is the number of groups (source: Esteban et al., 2012).

Polarization. This index is constructed as: $\Sigma_{i=1}^{m}\Sigma_{j=1}^{m} n_i^2 n_j k_{ij}$, where $k_{ij} = 1 - s_{ij}^{0.5}$ and s_{ij} is the degree of similarity between two languages, i and j, given the ratio of the number of common branches to the maximum possible number of 15 for the entire tree. n_i is the population share of group i and m is the number of groups (source: Esteban et al., 2012).

Greenberg Gini index. This index is constructed as: $G = \Sigma_{i=1}^{m}\Sigma_{j=1}^{m} n_i n_j k_{ij}$, where $k_{ij} = 1 - s_{ij}^{0.5}$ and s_{ij} is the degree of similarity between two languages, i and j, given by the ratio of the number of common branches to the maximum possible number of 15 for the entire tree. n_i is the population share of group i and m is the number of groups (source: Esteban et al., 2012).

Mountainous terrain. Proportion area of a country that is mountainous (source: Fearon and Laitin, 2003).

Non-contiguous. Noncontiguous states refer to countries with territory holding of at least 10,000 people and are separated from the land area containing the capital city either by land or by 100 kilometers of water (source: Fearon and Laitin, 2003).

Variation in elevation. This measure indicates standard deviation of elevation within country (source: Michalopoulos, 2012).

Mean land quality (soil-based). This measure presents average agricultural suitability of land across regions based on soil properties (source: Michalopoulos, 2012).