Population Surges and Social Conflict*

Daron Acemoglu† Leopoldo Fergusson‡ Simon Johnson§

April 12, 2011

Abstract

This paper examines the effect of population growth on violent social conflict. Exploiting the international epidemiological transition that began in the 1940s, we construct an instrument for changes in population (Acemoglu and Johnson, 2007) and find that countries with higher (exogenous) increases in population experimented larger increases in social conflict. A simple falsification test indicates that changes in conflict from 1900 to 1940 are uncorrelated with our instrument, lending support to our identification strategy. Our results are robust to using a variety of standard measures for conflict and are not driven by differential trends between countries with different baseline characteristics often emphasized in the literature on civil wars. Using a simple theoretical framework, we interpret these findings as an indication that a larger population, without a corresponding increase in resources and technology, generates greater competition for resources and makes conflict and civil wars more likely in environments where the institutions cannot handle the higher level of disputes.

PRELIMINARY AND INCOMPLETE. COMMENTS WELCOME. PLEASE DO NOT DISTRIBUTE WITHOUT AUTHORS' PERMISSION.

*Aknowledgements: MIT development lunch, Chicago, London, Universidad de los Andes, Universidad del Rosario. Ioannis Tokatlidis.

†Department of Economics, Massachusetts Institute of Technology, E-mail: daron@mit.edu

‡Department of Economics, Massachusetts Institute of Technology, E-mail: leopoldo@mit.edu

§Sloan School of Management, Massachusetts Institute of Technology, E-mail: sjohnson@mit.edu

1 Introduction

Civil wars have been one of the most major social events of the past several decades, creating enormous economic costs and human suffering in many countries. Standard measures indicate that, since 1960, internal violent conflicts of lower intensity have affected half of all nations in the world, and a third of all countries have suffered more intense civil wars (Blattman and Miguel, 2010).

This paper investigates the effect of population growth on violent conflict. Obviously, population growth and armed conflict can be jointly determined. There may be reverse causation, for instance if only peaceful countries can sustain population growth. Also, omitted variables may cause both population and conflict, and without properly controlling for such a variable we may spuriously estimate a causal effect of population on conflict when there is none. To solve these problems, we follow Acemoglu and Johnson (2007) in using the *international epidemiological transition* (the large improvements in life expectancy driven by international health interventions, more effective public health measures and the introduction of new chemicals and drugs that began in the 1940s).

This "natural experiment" provides us with an empirical strategy to isolate potentially-exogenous changes in population. In particular, the effects of the international epidemiological transition on a country's life expectancy and population were related to the extent to which its population was initially (circa 1940) affected by various specific diseases. Acemoglu and Johnson (2007) construct an instrument for population growth based on this pre-intervention distribution of mortality from various diseases around the world, and the dates of global intervention. The only source of variation in this variable, referred to as the "predicted mortality" instrument, comes from the interaction of baseline cross-country disease prevalence with global intervention dates for specific diseases. Since these innovations were exogenous to the conditions of each particular country, and since the instrument does not rely on the efficacy of each country's application of global programs, it provides the exogenous source of variation in population growth we are interested in.

Using this strategy, we find that countries with higher (exogenous) increases in population experimented higher increases in social conflict. Using various definitions of civil war and conflict commonly used in the literature, the instrumented changes in population have a significant and positive effect on the share of years per decade in which a country experiences civil war or conflict. The magnitude of our estimates indicate that a rise in log population of about 0.65 from 1940 to 1980, corresponding to the average change in (log) population in our sample of countries, causes roughly 4.3 additional years with civil war in 1980 relative to 1940. When considering lower intensity conflicts, the corresponding predicted effect is similar, of about 4.1 more years in conflict in 1980.

Our sample includes 65 countries listed in Appendix Table 1. Of these, 47 countries have complete data from 1940 or earlier (and include ten countries in Asia, 17 in Europe, 16 in Latin America, Australia and New Zealand, and Canada and the US). The remaining 18 countries, for which we have data from 1950 or 1960, include 12 non-Eastern European countries (four in Africa, six in Asia, two in Latin America) and six Eastern European countries. Thus, our sample has a good coverage of most regions in the world, with the exception of Africa. Of course, given the prevalence of violent civil conflict in Africa, this is an important limitation, but one that we cannot overcome with the available data. In particular, most African countries lack reliable data on causes of death disaggregated by disease dating back to the 1940s, and this is essential for our identification strategy.

Our results are robust to a variety of checks. Regressions using an alternative instrument show that our results are not driven by the choice for dating global interventions. In the alternative instrument, each country's initial mortality rate is assumed to decrease at the pace of the global mortality rate for the disease in question. Hence, this instrument is independent of the coding of global interventions.

Moreover, we verify that the connection between the predicted mortality instrument and conflict is not driven by a general long-run trend. More specifically, we conduct a simple falsification test using information on conflict prior to the epidemiological transition, from 1900 to 1940. Using this information, we find that while the change in our instrument from 1940 to 1980 is correlated with changes in conflict over the same period, it has no predictive power on the changes over the prior period. This simple "placebo" test gives additional support to the approach, by ruling out the possibility that these two variables are spuriously correlated because they follow a more fundamental long-run trend caused by a third factor.

Additional checks verify whether the effect of predicted mortality works not through changes in population, but through changes in the age composition of the population. The evidence suggests this is not the case. First, our set of diseases includes diseases that affect both adults and children. Second, there is no significant correlation in our sample between age composition and conflict. And third, by 1980 or after, when most of the change in conflict is observed, any changes in the demographic structure induced by the international epidemiological transition are much reduced.

Since civil wars are hard to measure, we make sure that our results are robust to using several different measures of internal armed conflict. The "Correlates of War" dataset, that covers a longer period than other sources and allows us to conduct our basic placebo test, constitutes our baseline source. We also use the backdated version of the Uppsala Conflict Data Project put together by the International Peace Research Institute. This alternative provides a lower threshold of annual deaths for inclusion, verifying that our results are present with low-intensity conflicts and not only limited to major civil wars. We also verify that our results hold when following Fearon and Laitin's (2003) suggestion of coding anticolonial wars as occurring within the empire in question. Also importantly, we verify that other related variables, such as state failure from the Polity IV State Failure Task Force, similarly increase for countries with higher exogenous increases in population. Moreover, our results are present when examining both the simple presence or absence of civil war, and also the intensity of conflict as measured by available estimates of battle-related deaths.

The evidence also indicates that our results are not spuriously identifying a relationship between population and conflict that is truly explained by differential trends (between countries with high and low population growth) caused by an omitted variable. The results hold when controlling for differential trends, parametrized as functions of various baseline country characteristics. For instance, since our results may in part be about wars that followed independence, we allow for differential trends depending on whether countries were already independent or not in 1940, finding similar results. Other baseline characteristics include the average quality of institutions, initial population and initial GDP. We do a similar check with variables that have often been emphasized in the literature on conflict, such as ethnolinguistic fractionalization, religion, and availability of natural resources.

The 1940s were of course a decade of big wars. To take this into account we not only examine the implications of population growth on conflict by examining panel regressions that exploit decade by decade variation changes in population and its implications on civil war. We also verify that our results hold when excluding the countries demographically most affected by World War II. In an additional exercise, we asssign the level of conflict of the 1950s to the 1940s. Finally, we run specifications ignoring the World War II years.

Since most available measures of civil wars used by scholars rely on absolute thresholds of number of battle deaths to categorize violence as civil war or conflict, greater population may mechanically increase the number of "detected" civil wars. To test whether this may partly explain our findings, we not only use several measures as explained above. We also check if, above what is predicted from the initial size of each country's populations, conflict increases more where population increased more. Also, we create a new definition of war, based on a relative threshold of violence.

That our results are largely robust to this set of specification checks lends further support to our conclusion that countries with higher (exogenous) increases in population after World War II experienced higher increases in social conflict. This conclusion is in line with theories dating back at least to Malthus, suggesting that large populations may affect social conflict. Along these lines, we use a simple theoretical framework to interpret our findings as suggestive that larger populations, without a corresponding increase in resources and technology, exacerbate the competition for resources and increase the likelihood of conflict and civil wars in environments where the institutions cannot handle the higher level of disputes.

Our results are of first-order importance for several reasons. First, there has been a tremendous increase in population in developing countries (the most common stage for internal warfare) in the last 100 years. Figure 1, depicting population and the logarithm of population for our sample of countries, reveals that those countries that were rich around the 1940s have experienced less rapid growth than initially-poor or middle-income countries. If greater population or sudden increases in population increase conflict, then post-World War II changes in population (partly a by-product of the international epidemiological transition) could have sizable effects on the likelihood of civil wars. This is an important additional dimension of the implications of population changes in less-developed countries.

A few examples convey the extent of these "population surges" that followed the international epidemiological transition, especially for initially poor countries. Ecuador, for instance, more than tripled its population from 1940 to 1980, moving from being a country of roughly 2.5 million people to having nearly eight million inhabitants. While sizeable, its population increase from 1900 to 1940 was much smaller, with an increase of roughly 75% (from a level of 1.4 million in 1900). This was not an isolated case. Other countries in our sample that roughly tripled their population from 1940 to 1980 include El Salvador (1.6 million to 4.5 million), Honduras (from 1.1 million to 3.6 million), the Phillipines (from 16.6 million to 51 million), and Thailand (15 million to 47 million). Also, where we have data, these rates of growth are significantly faster than those experienced by these countries during the first four decades of the century.

Contrast these increases with the increase in US population following the much discussed "baby boom" after 1945. In the US, population increased "only" around 70% from 1940 to 1980, starting at a level of about 130 million in 1940 and reaching roughly 228 million in 1980. Perhaps even more starkly, contrast the growth of these countries with the increase in population pressure over land in the United Kingdom from around the 1500s to the

1800s, a topic much discussed among economic historians given its potential repercutions for enclosures and the Industrial Revolution (see, for example, the review in Turner, 1984). According to Maddison, population in the UK rose from 3.94 million around 1500 to 6.17 million in 1600, 8.6 million in 1700, and finally reaching 21 million by 1820. In three centuries, this was a five-fold increase, and looking at the 120 years leading up to 1820 population increased by a factor of 2.2. Over a bit over two centuries, from 1700 to 1820, the UK population was multiplied by 3.4. In sum, many of the low income countries in our sample had a population increase in a couple of decades similar to what the UK had in a century or more.

In line with the above, a second important implication of the effect of population growth on civil conflict is that it may help explain a puzzling fact. As Figure 2 shows, the international epidemiological transition produced significant convergence in health conditions around the world. By the year 2000, the gap in average life expectancy at birth between initially rich and initially poor countries was reduced to about a half of its 1930 level, measured in absolute terms. However, in spite of the consensus that improving health can have large indirect payoffs through accelerating economic growth, no such convergence is apparent when examining output per capita. This is shown in Figure 3, which illustrates that while average (log) GDP per capita for initially poor, middle-income, and rich countries has trended upwards since the 1930s, poorer countries have not been able to outgrow and "catch-up" richer countries. The trends shown in these figures are robust to a more careful statistical analysis. When using the exogenous variation in health induced by the international epidemiological transition, Acemoglu and Johnson (2007) find little increases in GDP and a significant negative impact on GDP per capita. If increases in population exacerbated social conflict, the resulting productivity costs may help explain this result. Indeed, the relatively poorer countries experienced greater improvements in health and population. But they also experienced the largest increases in civil wars, as depicted in Figures 4 to 7 using the various measures of conflict that will be considered in this paper.

Our paper is related to several strands of research. A large line of research relies on cross-country evidence to study the causes of civil war. Following contributions such as those of Collier and Hoeffler (1998; 2004) and Fearon and Laitin (2003), scholars have emphasized poverty, inequality, weak institutions, political grievances, and ethnic divisions as explanations for the outbreak and persistence of civil war. Yet, the validity of the statistical inference exercise challenges many of these studies. With a few exceptions[1], they do not adequately address the possibility of reverse causality, or omitted variables bias, driving observed correlations. Indeed, Blattman and Miguel (2010) conclude in their survey of the literature that "further cross-country regressions will only be useful if they distinguish between competing explanations using more credible econometric methods for establishing causality" (p. 8).

To the best of our knowledge, Brückner's (2010) paper is the only previous work that addresses these causality issues in a study of population size and civil conflict[2]. It studies a panel of 37 Sub-Saharan countries during the period 1981–2004 and uses randomly occurring droughts as an instrumental variable for population. In line with our findings for a sample of non-African countries, Brückner finds that (instrumented) population size has an economically meaningful and statistically significant effect on African civil conflict. The most important threat to the validity of his estimates is that droughts affect conflict through their effect on other variables, such as income, and not via population. To tackle this issue, Brückner (2010) estimates the impact of population on civil conflict risk conditional on per capita GDP, and uses rainfall and international commodity prices as additional instrumental variables to deal with the endogeneity of income. While, as noted above, we are unable to use our identification strategy for most African countries given lack of reliable disaggregated disease data dating back to the 1940s, Bruckner's results are very suggestive that the main implications of population surges studied in our sample apply to Africa also.

Somewhat surprisingly, population has not been a prime focus in the theoretical literature on the "economics of conflict" (see, for example, the survey by Garfinkel and Skaperdas, 2007). However, there has been a lively debate on the effects of population pressure on violent conflict in other disciplines, in particular political science[3]. Broadly (and simplifying) the debate is between one that highlights "Malthusian" channels to conflict and other

[1] Exceptions include Miguel, Shanker Satyanath and Sergenti (2004), who use annual rainfall growth as an instrument for income growth in sub-Saharan Africa; Besley and Persson (2008), who rely on plausibly exogenous international commodity price movements; and Brückner (2010), which we review below.

[2] In most of the empirical economics literature on conflict, population is a control variable (often with a positive sign), but it is rarely the prime focus and there is no attempt to control for its endogeneity. For instance, in Sambanis (2002) review of this research, the role of population is hardly mentioned. Collier and Hoeffler (2004) report a positive coefficient on population, which the authors interpret as consistent with either a greed or grievance story for conflict, but their regressions for a panel of countries do not control for country fixed-effects and thus may well be driven by omitted country-specific characteristics. Indeed, in Fearon and Laitin's (2003) study of conflict onset, the positive coefficient on population disappears once fixed effects are included in the regression. Miguel, Satyanath, and Sergenti (2004) also report a positive coefficient, but their focus is on the effect of income on conflict.

[3] The connection between population and conflict has also received significant public attention, as testified by Robert Kaplan's famous 1994 essay "The Coming Anarchy," in turn heavily influenced by Homer-Dixon.

which is more skeptical about the practical relevance of such arguments. Among the former, Homer-Dixon (1991, 1999) studies the connection between population growth, pressure on environmental resources, and conflict. For Homer-Dixon, poor countries are in general more vulnerable to environmentally-induced conflicts. Moreover, while recognizing (as do many anti-Malthusians) resilience and adaptability in human-environmental systems, he contends that "as population grows and environmental damage progresses, policymakers will have less and less capacity to intervene to keep this damage from producing serious social disruption, including conflict" (Homer-Dixon, 1991). During the 1990s, Homer-Dixon and his collaborators on the Project on Environment, Population and Security (EPS), often referred to as the Toronto Group, produced a number of case studies on "environmental conflict" which, though not concerned solely with population growth, did underscore its importance as potential source of environmental scarcity and consequently conflict (Homer-Dixon, 1994).

The overall approach of these studies on "environmental security" has come under attack from a number of researchers. Richards (1996) dismisses what he calls the "New Barbarism" theory of conflict put forward by Kaplan's (1994) essay. In particular, for the case of Sierra Leone, he refutes the neo-Malthusian "believe that ordinary Africans will 'sink their ship' by over-production, unless checked by famine, war and disease [and that] war is a process through which the poor in Africa will learn its limits" (p. 121). Instead, he notes that the process of forest conversion in Sierra Leone has taken place over many centuries, and that local land-users have responded in sensible way to its different phases, with no evidence of environmental degradation spiralling out of control prior or around the years of civil war between the government and the Revolutionary United Front. He concludes that in Sierra Leone "war is a consequence of political collapse and state recession, not environmental pressure (...) Violence has been incubated in forest fastnesses. [The problem] is too much forest, not too little" (p. 124).

Gleditsch (1998) also offers a critique of this line of research, but emphasizes its methodological shortcomings. Most importantly, he argues that much of it fails to qualify as "systematic" quantitative or comparative research, and violates the rules of quasi-experimental methodology[4]. On the specific issue of population density (or population growth) and vi-

[4] His full critique includes 9 more specific "problems" of the literature: 1. There is a lack of clarity over what is meant by "environmental conflict"; 2. Researchers engage in definitional and polemical exercises rather than analysis; 3. Important variables are neglected, notably political and economic factors, which have a strong influence on conflict and mediate the influence of resources and environmental factors; 4. Some models become so large and complex that they are virtually untestable; 5. Cases are selected on values of the dependent variable; 6. The causality of the relationship is reversed; 7. Postulated events in the future are cited as empirical evidence; 8. Studies fail to distinguish between foreign and domestic conflict; and, 9.

olence, he notes that while "strictly speaking, these are not measures of either resource scarcity or environmental degradation," (p. 384) they may provide an indirect measure and, at least in this area, a few papers have provided systematic cross-national evidence. The evidence has not been conclusive[5]. More importantly, these studies typicaly do not address the endogeneity of population in regressions for conflict.

The paper proceeds as follows. In section 2 we present a simple motivating theory capturing the "Malthusian" mechanisms that may lead from population growth to conflict. Section 3 describes our data, and Section 4 presents ordinary least square (OLS) results. Section 5 discusses our identification strategy, and Section 6 shows our main results from two-stage least squares (2SLS) estimates. Section 7 presents a series of robustness checks on our estimates, and Section 8 is a more speculative section on possible consequences of population growth for regime changes. Section 9 concludes.

2 Motivating Malthusian Theory

In this section we present a simple framework capturing the "Malthusian" mechanisms that may lead from population growth to conflict. The basic idea is that population growth generates greater rents to a fixed factor relative to labor, and this form of "scarcity" makes conflict more likely. In the specific case of land as the fixed factor, this mechanism is particularly relevant in less-developed economies, where agriculture and traditional activities matter most.

It must be noted that the connection between population growth and conflict is not a necessary one. Indeed, it is not necessarily true with constant returns to scale to variable factors. However, we show that when greater population increases "scarcity," it also makes conflict more likely.

Confusion reigns about the appropriate level of analysis.

[5] In particular, Tir and Diehl (1998) examine the Correlates of War dataset to evaluate the impact of population growth and density on international conflict involvement, initiation, and escalation over the period 1930-89. They find that population growth pressures have a significant impact on military conflict involvement, especially in poor countries, but no correlation with conflict initiation or escalation, or between population density and conflict. Hauge and Ellingsen (1998) find that factors like deforestation, land degradation, and scarce supply of freshwater, alone and in combination with high population density, increase the risk of domestic armed conflict, especially low-level conflict, in the period 1980-92. However, economic and political variables prove more decisive than environmental scarcity in predicting the incidence of domestic armed conflict. For more studies along these lines, see the special 1998 issue of the Journal of Peace Research and Diehl and Gleditsch (2001).

Suppose that aggregate output is given by a constant returns to scale production function of a fixed factor, Z, and labor, N, and is also a function of technology, A:

$$Y = F(Z, N, A) \equiv f(N), \tag{1}$$

where $F(\cdot)$ exhibits constant returns to scale in (Z, N) and f gives output as a function of labor, holding technology and Z constant. Thus, if N increases with A constant, output per worker, $f(N)/N$, declines. However, if increases in labor (population) come together with increases in the technology parameter A, output per worker can remain constant, thus avoiding scarcity.

We assume the following simple allocation of resources. Each individual i in society supplies one unit of labor inelastically and also owns a fraction θ_i of land. For simplicity, we also suppose markets are competitive, though this is not important for our analysis. With these assumptions, individual income and consumption is given by

$$c_i(N, \theta_i) = f'(N) + \theta_i [f(N) - Nf'(N)]. \tag{2}$$

The key observation from equation (2) is that the marginal increase in an individual's consumption from an increase in his landholdings, is larger when population increases,

$$\frac{\partial^2 c_i}{\partial N \partial \theta_i} = -Nf''(N) > 0.$$

That is, land shares matter more for consumption when population is greater. The intuition is simple: with higher N, land rents are more important relative to wages due to diminishing returns. This implies a *Malthusian channel to conflict* when land can be contested with violence.

To explore this channel, imagine the society consists of two groups, 1 and 2. All members within a group are identical, and also, to simplify the discussion we suppose both groups are of size $N/2$ and population growth leaves relative shares unchanged. To capture the disruption costs of conflict, assume that if a group initiates conflict, then this reduces output to a fraction $(1-\rho)$ of what it would have been without conflict. Group j has probability p_j of winning the conflict and if so, it captures a fraction λ_{-j} of the land of the other group (λ loosely an inverse measure of the "specificity of assets"). With probability $p_{-j} = 1 - p_j$, it loses the conflict and a fraction λ_j of its land. Also for simplicity, the advantage of being the first mover and deaths from conflict are ignored. Also, as discussed below, voluntary concessions to avoid civil war are ignored. Finally, assume that all agents are risk neutral. Then, the expected benefits to conflict, $\pi_j(N,\theta,\lambda,\rho)$, for group j is given by

$$\begin{aligned}\pi_j(N,\theta,\lambda,\rho) &= -\rho\left\{F'(N)+\theta_j\left[f(N)-Nf'(N)\right]\right\} \\ &\quad +(1-\rho)\left[p_j\lambda_{-j}\theta_{-j}-p_{-j}\lambda_j\theta_j\right]\left[f(N)-Nf'(N)\right].\end{aligned} \tag{3}$$

In (3),the first line captures the destruction costs of conflict. The second line captures potential benefits, amounting to the undestroyed expected additional land rents that will be expropriated with violence. For there to exist equilibrium conflict, a necessary (but not sufficient condition) is to take a society in which

$$p_j\lambda_{-j}\theta_{-j}-p_{-j}\lambda_j\theta_j \neq 0.$$

so that one of the groups will have potential gains from conflict—say group j. These gains are likely to be higher when θ_{-j} and λ_{-j} are high, so greater inequality of resources between the two groups and lower degree of specificity will contribute to the potential gains from conflict. But even in this case $\pi_j(N,\theta,\lambda,\rho)<0$ is possible for both groups because of the first term—*cost of disruption.*

The same reasoning as above in our discussion of equation (2) implies that whenever $\pi_j(N,\theta,\lambda,\rho)=0$,

$$\frac{\partial \pi_j(N,\theta,\lambda,\rho)}{\partial N}>0.$$

Therefore, an increase in population makes the group that is more likely to initiate civil war more "pro civil war." As noted before, this result does not apply when N increases in tandem with A. This observation is important, in the sense that the Malthusian mechanism says nothing about secular increases in population. Rather, the predictions are about the level of population for given A or for "unusual increases in population".

This simple framework generates other intuitive comparative static results. Greater inequality (θ) between groups makes conflict more likely. Lower disruption costs (lower ρ) and lower asset specificity (higher λ), makes conflict more likely. The latter results may be linked to the importance of natural resources and agriculture versus human capital and industry. In particular, modern capitalism depends on production processes that may be easily disrupted with violence. When traditional production methods are prevalent, for instance when the main form of capital is land, the costs of violence are relatively smaller. Presumably, the productivity of land is harder to destroy than the productivity of a factory. Also, in less-developed economies there is less investment in specific assets, such as human capital, which not only cannot be expropriated through violence but which, unlike land, can move to other places with the outbreak of violence (Acemoglu and Robinson, 2006).

A central question that we have ignored is: Why is conflict not prevented by more efficient ways of redistributing resources. Why not make voluntary concessions and avoid the cost ρ of conflict? Commitment problems are a prominent explanation (Acemoglu and Robinson, 2001, 2006; Fearon, 1998, 2004; Powell, 2006; Acemoglu, Egorov and Sonin, 2008). To see this, consider the same environment in a dynamic setting, but in each period there is a probability $q < 1$ that either group can initiate civil war. Assume all agents have discount factor $\beta \in (0, 1)$. To simplify the discussion, assume that, after civil war, there is a permanent redistribution of resources and never any social conflict again. Also to simplify, only cash transfers are feasible and group 1 is the one considering civil war.

In this context, the benefits from civil war for group 1 are proportional to $1/(1-\beta)$—because of discounting. If the group is sufficiently patient (β is high enough), then cash transfers in a given period are not sufficient to offset this gain. But group 2 cannot make a credible promise to make the cash transfers in the future once the window of opportunity for civil war disappears. In this setting, civil wars arise along the equilibrium path even though more efficient ways of dealing with conflict exist. In particular, fix $\beta \in (0, 1)$, then there exists $\bar{q}$ such that for all $q < \bar{q}$, the Markov Perfect Equilibrium will involve equilibrium civil war. Also, there exists $\hat{q} < \bar{q}$, so that for all $q < \hat{q}$, all Subgame Perfect Equilibria involve civil war.

3 Data[6]

Measuring civil war and conflict is not simple. The most commonly used datasets differ in the data sources they draw from and in important aspects of the coding procedure, such as thresholds of violence required, definitions of war onset and termination, and treatment of outside parties, communal violence or state repression (see Sambanis, 2004). For this reason, we rely on a number of alternative databases to measure conflict.

We measure conflict as the ratio of number of years in conflict to total years for a period around a reference date t (where, typically, $t = 1940, 1950, ..., 1980$). Thus, we focus on conflict incidence, rather than onset, per decade. This is a natural choice since we are

[6] Here we describe our main dependent and independent variables, and a full description of all variables and sources including all controls and baseline characteristics can be found in Appendix Table 1.

focusing on a relatively long-term phenomenon: increases in population over a period of several decades and the potential response of conflict. Relatedly, since datasets may disagree on the exact year of conflict initiation, the differences between them is typically larger for conflict onset than for conflict incidence. This problem must be addressed when examining high-frequency responses of conflict, but in our case it is better to select the most robust measure of conflict.

Our baseline dataset for conflict is version 3 of the Correlates of War (henceforth COW) dataset (Sarkees, 2000). This source records civil wars since 1816, which allows us to run a simple falsification test using pre-existing trends in conflict. A civil war is defined as a war fought within state borders, between a government and non-government forces, where the central government is actively involved in military action, with effective resistance for both sides[7], and with at least 1000 battle-related deaths during the war. This is a higher threshold of violence for inclusion than alternative sources, which we also use for robustness. When using COW, we assign the number of years with conflict to the reference dates as follows: wars from 1940-1949 are assigned to 1940, wars from 1950-1959 are assigned to 1950, etc.

Our second database for conflict, available since 1946, is the backdated version of Uppsala Conflict Data Project, in conjunction with International Peace Research Institute (UCDP/PRIO Armed Conflict Dataset Version 4, Gleditsch et al, 2002). We assign number of years in conflict to reference dates as follows: 1946-1949 to 1940, 1950-1959 to 1950, 1960-1969 to 1969, etc. Of course, in the case of 1940, we divide the number of years in war by 4, and not by 10 like in the other reference years. This dataset includes conflicts where at least one of the primary parties is the government of a state, and where the use of armed force results in at least 25 battle-related deaths per year. The dataset includes four types of conflicts, and we use the two categories for internal conflict ("internal armed conflict" and "internationalized internal armed conflict").

As a third database to verify the robustness of our results, we consider Fearon and Laitin's (2003) coding of civil war. This dataset covers the period 1945-1999, and the criteria are broadly similar to those of COW[8], except that anticolonial wars are coded as occurring within the empire in question (e.g. Algeria is assigned to France). As with the other datasets, we

[7] To constitute effective resistance, both sides must have been initially organized for violent conflict, or the weaker side must be able to inflict the opponents at least five percent of the number of fatalities it sustains.

[8] Conflicts are included if they: involved fighting between agents of (or claimants to) a state and organized, nonstate groups who sought control of a government, region, or change in government policies; killed at least 1,000 over its course, with a yearly average of at least 100; at least 100 were killed on both sides (including civilians attacked by rebels).

count the number of years with any incidence of war, and use a similar assignment to reference dates ($1940 = 1945 - 1949$, $1950 = 1950 - 1959$, etc.).

To examine effects on the intensity of conflict and as further robustness, we use information on battle deaths from the Center for the Study of Civil War (CSCW) Battle Deaths Dataset (Lacina and Gleditsch, 2005). We use version 2, compatible with the COW dataset, and use the "best estimate" of annual battle-related deaths (similarly assigning deaths to reference years using the rule: $1940 = 1940 - 1949$, $1950 = 1950 - 1959$, etc.). We also use version 3, compatible with the UCDP/PRIO dataset, since this dataset has a smaller threshold of battle-deaths for inclusion. This allows us to more convincingly check the robustness of our results to possible mechanical effects on detected wars with increases in population.

We also explore whether the large exogenous increases in population in our data have an effect on state failure and regime changes. As a measure of state failure, we use the "extreme political failure" dummy, available from 1955 from the Political Instability Task Force (Phase III), with similar assignment to dates ($1950 = 1955 - 1959, \ldots$). For regime changes and democratic changes we use Polity IV.

We have data for 65 countries listed in Appendix Table 1, although only for 47 of them we have complete data from 1940 or earlier. The remaining 18 countries, for which we have data from 1950 or 1960, include 12 non-Eastern European and six Eastern European countries. As highlighted previously, we are able to include only four African countries (Algeria, Egypt, South Africa, and Tunisia), an important limitation given the prevalence of Civil War in Africa. However, lack of reliable data on causes of death disaggregated by disease dating back to the 1940s explains the exclusion of most of Africa in Acemoglu and Johnson (2007). Our efforts to extend that sample proved futile.

The treatment of countries which enter or leave the state system or whose borders change over time must be explained. For each country, we check when the respective datasets considers the country as entering or leaving the state system, and adjust accordingly the number of years on a reference year. Thus, for example, if Algeria enters the COW system membership in 1962, the measure of conflict for 1960 is the number of years in conflict from 1962-1969 (if any), divided by 8 (instead of 10). Moreover, we code as missing (not zero) the observations for Algeria in reference years prior to 1960. While this surely is a reasonable choice (Algeria did not exist as a state before, just like Bangladesh did not prior to 1971, and so on), the criteria for inclusion in the COW system membership include a population threshold of 500,000 and having diplomatic recognition (prior to 1920, recognition at or above the rank of *charge d'affaires* with Britain and France and, later, being a member of the League of Nations or the United Nations, or receiving diplomatic missions from two major powers). Thus, for instance, Costa Rica and Australia are not in the system in 1900. While it may be seem reasonable to include them as (peaceful) states in 1900 for our falsification regressions, we avoided doing any of these adjustments to the data, instead following the approach of abiding by the choices made by the authors of each of the independent codings of civil war.

As a general rule, when a country splits we do not attempt to code "external" wars between formerly member states as "internal" wars in the territory as a whole. Again, we follow this procedure to avoid including our own criteria of what constitutes an internal conflict, instead abiding by the classification of the various sources we use, and verifying the validity of our results to dropping potentially problematic countries or territories.

Also, we adopt the general rule of aggregating civil wars and regime changes for the splitting territories. The countries in our sample potentially affected by this choice are just Czechoslovakia, Germany, the USSR and Vietnam. Thus, for example, we add USSR internal conflicts while it existed, and aggregate *internal* conflicts (if any) and regime changes of the member states and assign them to the USSR as a whole after 1991. As it turns out, choosing this against other alternatives makes little difference in practice for the coding of our dependent variables. First, our main specifications end in 1980, prior to many of these splits. Second, in many cases the dependent variable would be the same aggregating the territories or not. For instance, in Czech Republic in the 90s, our dependent conflict variables are always zero with or without aggregating Slovakia. Also, we would have 1 in our regime transition dummy in the 1990s whether we took Czech Republic, Slovakia, or the aggregate.

This procedure also minimizes potential mismatches between the level of aggregation of the population figures from Maddison (2006) and civil conflict/political data. Indeed, in the case of Czechoslovakia/Czech Republic, Maddison presents throughout the data for Czechoslovakia as a whole, even after split between Slovakia and the Czech Republic. Similarly, population figures are for Vietnam as a whole, and for the USSR while it existed and later the total for ex-USSR. Maddison's treatment of Germany is more complicated. It takes the 1870 frontiers until 1918, the 1936 frontiers for 1919-1945, and present frontiers from then on. Also, it must be noted that the immediate post-war disease data from the UN are divided into Eastern Germany, Federal Republic of Germany, Berlin and West Berlin, and numbers for the Federal Republic were used in Acemoglu and Johnson (2007). To make sure our results do not depend on any of these choices, we also dropped Czechoslovakia, Germany, the USSR and Vietnam and found results similar to those reported below.

The sources to build our instruments can be found in Acemoglu and Johnson (2007)[9]. Information on age structure is from the United Nations. We also consider a number of controls for robustness exercises, all of which are described in Appendix Table 1. These include measures of institutions, whether our countries were independent in 1940 or not, whether the country was affected by World War II, initial (in 1930) GDP per capita, availability of natural resources (diamonds, oil, and gas), ethnic and religious fragmentation and fractionalization, and share of Catholic, Muslim, and Protestant populations.

Table 1 presents descriptive statistics (sample means and standard deviations) for our main data. We present these summary statistics for the sample as a whole, for groups of countries by income, as well as dividing them between countries experiencing a change in predicted mortality above and below the median. Column 2 shows a general trend, evident in every measure of conflict, of increasing conflict from 1940 to 1980. Also, columns 3 to 5 show that such an increase is concentrated in middle-income and, especially, poor countries. But more importantly, comparing the change in our conflict measures from 1940 to 1980 in columns 6 and 7, we observe that countries above median change in predicted mortality exhibit larger increases in conflict than those below the median change. For instance, the average years in conflict (per decade) according to the COW measure increased from 0.8 years to 2.5 years for countries with above median change in predicted mortality from 1940 to 1980, while it decreased from 0.2 years to 0.06 years for those with below-median change. These patterns are suggestive for our hypothesis, and we will examine below if they survive in our regression exercises and various robustness checks.

4 Ordinary Least Squares (OLS) Results

As a benchmark for our IV estimates reported below, and to examine more systematically these correlations in our data, we start by presenting results for simple ordinary least squares (OLS) regressions of our measures of conflict on population. More specifically, in Table 2 we report regressions of the form

$$c_{it} = \pi x_{it} + \zeta_i + \mu_t + \mathbf{Z}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \tag{4}$$

[9]The main source of the necessary health data on incidence of diseases circa 1940 is the League of Nations (based on national statistics), but other sources were consulted for consistency.

where c_{it} is one of our measures of conflict for country i and reference year t, and x_{it} is (logarithm of) population. ζ_i denote a full set of country fixed effects while μ_t are a full set of year dummies, which we always include to remove time-invariant country-specific factors and global trends affecting population and conflict. Z_{it} is a vector of other controls. In Table 2, as in subsequent tables, we present two types of estimations. Panel A presents *long difference* estimations in which just 2 dates, usually 1940 and 1980, are included in the regressions. In this case, (4) is equivalent to a regression of the change in conflict between the two dates on the change in population between the same two dates, giving us a particularly simple interpretation. In Panel B of the Table, we look at panel results between 1940 and 1980, with data for $t = 1940, 1950, 1960, 1970, 1980$.

For all of our regressions, we calculate standard errors that are fully robust against serial correlation at the country level (e.g., Wooldridge 2002, 275)[10]. However, these standard errors may be downward biased due to the relatively small number of clusters. Thus, we also implemented the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). In particular, we performed wild bootstraps resampling our clusters with replacement, and calculated bootstrap p-values for the t-statistics of our main coefficients of interest. Our regression tables report p-values following both of these approaches.

The OLS results in columns 1-3 of Table 2 reveal that population is positively correlated with conflict. The estimated coefficient for log population (0.303) in the long-difference regression in column 1 of Panel A, where conflict is measured using the COW dataset, implies that the average change in log population in our sample of 0.65 is correlated with about 1.97 more years in conflict in 1980 relative to 1940. For a country like El Salvador, with an increase in population from 1.6 to 4.6 million, this predicts about 3.2 more years in conflict during the 1980s than during the 1940s. The size of the coefficient is fairly stable across different datasets for conflict, as columns 2 and 3, which use the UCDP/PRIO and Fearon and Laitin datasets, reveal. And while these latter estimates are not significant at conventional levels when using the standard cluster-robust p-values, we find reassuringly smaller p-values (of around 8% to 10%) with the wild bootstrapping procedure. In fact, this is a general pattern throughout this and other tables, suggesting that the degrees of freedom adjustment on the asymptotic cluster-robust covariance matrix is in fact too conservative in our case.

[10] Moreover, as a "degrees of freedom adjustment" we adopt the relatively conservative formula for the covariance matrix $\frac{N-1}{N-G-k}\frac{M}{M-1}V$ where N is the number of observations, k the number of coefficients (excluding the fixed effects and constant), M the number of clusters, and V is the asymptotic cluster-robust covariance matrix. A less demanding alternative often used is $\frac{N-1}{N-k-1}\frac{M}{M-1}V$.

Also interestingly, column 4 shows a strong and very significant population between population and state failure. The estimated coefficient for battle deaths as the dependent variable, in column 5, is also positive. However, the p-values reveals that it is, if anything, only marginally significant.

One possible concern with the results from Table 2 is that they might by driven by age composition effects. In particular, rather than *larger* populations it may be that *younger* populations are correlated with conflict. Table 3 examines this issue by presenting regressions as in Table 2 with the share of population from 15 to 34 years of age as the independent variable. As the odd columns from this Table reveal, this variable is usually not significantly correlated with conflict. Moreover, when population is also included as a regressor in the even columns of the Table, the results for population are very similar to those in Table 2, and the point estimate on the share of young population is usually negative. These results suggest that rather than age composition, total population is correlated in our sample with conflict (and state failure). However, these OLS estimates are not necessarily causal, and the true effect of population on conflict might be larger or smaller than implied by these coefficients. We next investigate this question.

5 Identification Strategy: the International Epidemiological Transition as a Natural Experiment

As noted in the Introduction, our identification strategy relies on the *International Epidemiological Transition* as a "natural experiment" creating large increases in population. Such increase in population followed major exogenous (to most countries) innovations in drugs (e.g., penicillin) and associated effective treatments, and chemicals (e.g., DDT) and how to apply them. International programs to spread best practices followed through, led by international agencies such as the WHO and UNICEF. This episode provides an instrument for population growth, by exploiting pre-intervention distribution of mortality from various diseases around the world, and the dates of major global interventions affecting mortality from this set of diseases.

More specifically, we use the *predicted mortality instrument*, which adds each country's initial (in 1940) mortality rate from 15 diseases until there is a global intervention, and after the global intervention, the mortality rate from the disease in question declines to the frontier mortality rate[11]. For country i at time t, the instrument is constructed as follows:

$$M_{it}^{I} = \sum_{d \in D} ((1 - I_{dt}) M_{di40} + I_{dt} M_{dFt}), \tag{5}$$

where: M_{di40} denotes mortality in 1940 (measured as number of deaths per 100 individuals per annum) for country i, from disease $d \in D$; I_{dt} is a dummy for intervention for disease d that takes the value of 1 for all dates after the intervention; M_{dFt} is mortality from disease d at the health frontier of the world at time t; and D is the set of diseases listed above. Since M_{di40} is the pre-intervention mortality for disease d and $I_{dt} = 1$ after a global intervention, the variation in this variable comes from the interaction of baseline cross-country disease prevalence with global intervention dates for specific diseases. Countries that experienced higher mortality than others for a given disease are expected to observe larger increases in population after the intervention.

One concern with the predicted mortality instrument is that it depends on the choice for dating global interventions. An alternative instrument that is independent of the coding of global interventions assumes each country's initial mortality rate decreases at the pace of the global mortality rate for the disease in question. The formula for this *global mortality instrument* is given by

$$M_{it}^{I} = \sum_{d \in D} \frac{M_{dt}}{M_{d40}} M_{di40}, \tag{6}$$

where M_{dt} (M_{d40}) is global mortality from disease d in year t (1940), calculated as the unweighted average across countries in the sample of countries in Acemoglu and Johnson (2007).

We use these variables as instruments for population. In particular, we posit the first-stage relationship for country i at time t,

$$x_{it} = \varphi M_{it}^{I} + \tilde{\zeta}_i + \tilde{\mu}_t + \mathbf{Z}_{it}' \tilde{\boldsymbol{\beta}} + u_{it} \tag{7}$$

where: x_{it} is the logarithm of population; M_{it}^{I} the predicted (or global) mortality instrument; $\tilde{\zeta}_i$ is a full set of country fixed effects; $\tilde{\mu}_t$ are year fixed effects; and $\mathbf{Z}_{it}$ represents a vector of other controls.

Acemoglu and Johnson (2007) show that the changes in predicted mortality led to major

[11] The 15 diseases are (in rough descending order of importance): malaria, pneumonia, and tuberculosis; influenza, cholera, typhoid, smallpox, shigella dysentery, whooping cough, measles (rubeola), dyphteria, scarlet fever, yellow fever, plague, typhus.

improvements in life expectancy (and other health measures). This is shown in Figure 8, where the horizontal axis depicts the change in predicted mortality between 1940 and 1980, and the vertical axis shows the change in log life expectancy during the same time period. In countries like India, Pakistan, Indonesia, Ecuador and El Salvador, where predicted mortality declined by a large amount, there were large gains in life expectancy. Instead, life expectancy remained largely unchanged in parts of western Europe, Uruguay, Argentina, Korea, and Australia, where predicted mortality did not decrease as much. Also importantly, Figure 9 depicts the same negative relationship holds without the richest countries, so it is not driven by the comparison of initially rich countries to initially low- and middle-income countries.

One concern is that these results are driven by a preexisting trend. To show that this is unlikely to be the case, Figures 10 (and 11) look at changes in life expectancy during the pre-period, 1900–1940 (and 1930-1940), and examine whether they correlate with future (post-1940) changes in predicted mortality. Unlike Figures 8 and 9, there is no evidence of a negative relationship. These issues are examined in greater detail in Acemoglu and Johnson (2007), and the results point to a very robust and significant relationship between predicted mortality and health that is unlikely to be driven by preexisting trends. We now build on the consequences of these improvements on health for population growth to examine the relationship between population and conflict.

6 Main Results

6.1 First Stages

That there is a strong first-stage relationship between predicted mortality and population is evident in Figures 12 and 13. Like Figures 8 and 9, these figures measure the change in predicted mortality between 1940 and 1980 in the horizontal axis, but show the change in log population during the same time period in the vertical axis. Again, we observe a strong negative relationship (whether looking at the whole sample or at low- and middle-income countries alone).

Table 4 shows the first-stage relationship in regression form by estimating equation (7). This Table shows the strong negative relationship between log population and predicted mortality is robust across alternative samples. Panel A reports long-difference specifications, and panel B reports panel regressions. Column 1 includes all countries in our sample, and shows an estimate of φ equal to -0.782, which is significant at less than 1 percent. This estimate implies that an improvement in predicted mortality of 0.469 per 100 (or 469 per $100,000$, which is the mean improvement between 1940 and 1980 in our base sample) leads to around a 44% percent increase in population (the mean population in our sample in 1940 was about 34 million, so this is an increase of nearly 15 million, whereas the actual mean increase in population between 1940 and 1980 was about 30 million). This implies that changes in predicted mortality account for almost one-half of the increase in population 1940 and 1980. Column 2 repeats the same regression excluding Eastern Europe, and Column 3 looks only at low- and middle-income countries. The estimate of φ is similar , and still significant at less than 1 percent. Panel B repeats the same regressions using a panel with decadal observations. The results are still highly significant but the coefficients smaller, which is reasonable since these regressions exploit shorter-run responses to changes in predicted mortality. Column 4 presents results using the global mortality instrument. The results are also strong and significant, reassuring us that they do not depend on the coding of global intervention dates. Finally, column 5 excludes the countries most affected by World War II, again with almost identical results.

The main potential threat to the exclusion restriction would be that the baseline mortality rates are correlated with future changes in conflict. To show that this is unlikely to be the case, we will show the robustness of our IV results to the inclusion of differential trends that are parametrized as functions of various baseline characteristics. Whether this explains the first-stage relationship is investigated with regressions of the form

$$x_{it} = \varphi M_{it}^{I} + \tilde{\zeta}_i + \tilde{\mu}_t + \sum_{t=1940}^{1980} \kappa_i' \bar{\omega}_t + u_{it} \qquad (8)$$

where $\bar{\omega}_t = 1$ in year t and zero otherwise, and κ_i are "time-invariant" characteristics of country i. These characteristics include: a measure of the average quality of institutions (average of the constraints on the executive from the Polity IV data set over 1950-70; a dummy for independence in 1940; initial (in 1930) GDP per capita and population; and measures of the availability of natural resources and ethnic polarization/fragmentation, often emphasized on the empirical literature on civil war. These regressions are reported in Table 5. Since equation (8) includes a full set of time interactions with c_i, differential trends related to these characteristics are taken out. In long-difference regressions of panel A, this specification is equivalent to including an interaction between the 1980 dummy and the baseline characteristics.

The results in both panels show that controlling for these characteristics has little effect on our results. Overall, the evidence in this section shows that the instrument is very strong and that its correlation with population is unlikely to be driven by differential trends due to a third factor.

6.2 Reduced Forms and Falsification

Figures 14 and 15 find a strong reduced-form relationship between the change in conflict 1940-1980 and the change in predicted mortality over the same period, for all countries in our sample and for low- and middle-income countries only, respectively. These patterns are shown in regression form in columns 1 and 2 of Table 6: countries with a larger decline in predicted mortality experienced a larger increase in years in conflict. Given the negative relationship between predicted mortality and population shown in the previous section, this translates into a positive effect of population on conflict in our 2SLS estimates below.

Before presenting these estimates, a useful falsification exercise (as we did before for the case of life expectancy) is to look at changes in predicted mortality, and see whether they correlate with changes in conflict or population during the pre-period. Columns 3 and 4 of Table 6 find no relationship between the change in conflict from 1900 to 1940 and change in predicted mortality from 1940 to 1980, for the base sample and for low- and middle-income countries. This lack of relationship can be seen in Figures 16 and 17. There is a strong contrast between the patterns in these figures and those of Figures 14 and 15. Similar specifications for our first stage (changes in log population from 1900 to 1940 and in predicted mortality from 1940 to 1980) are shown in columns 5 and 6, again with no sign of such a relationship. Predicted mortality explains changes in population after 1940, but not before 1940.

These results therefore suggest that there were no preexisting trends related to changes in predicted mortality either in population or in our key conflict outcome variables. This gives us greater confidence in using predicted mortality as an instrument to investigate the effect of population on conflict. However, the exclusion restriction for our IV strategy - $Cov(M_{it}^{I}, \varepsilon_{it}) = 0$, where ε_{it} is the error term in the second-stage equation- requires that the unique channel for casual effects of predicted mortality on conflict is changes in population. Since this exclusion restriction is fundamentally untestable, we cannot be entirely sure that this is the case. As we documented before, for example, changes in predicted mortality generated major changes in life expectancy. However, life expectancy should probably not have a direct effect on conflict aside from through its impact on population. As noted before when discussing Table 6 of first stages, as an additional check we will control for differential trends parametrized as functions of baseline characteristics in our 2SLS regressions to be discussed below.

As noted when discussing the OLS estimations, another concern is that the casual channel is partly age structure. While potentially important, the OLS results of Table 3 do not show a significant correlation. In addition, by 1980, when we observe the largest impacts on conflict, changes in the demographic structure induced by the international epidemiological transition are much reduced. Table 7 investigates this issue further. Column 1 presents long-difference regressions for our base sample (where $t = 1940$ and 1980) of the share of population of ages 15-34 on predicted mortality. The coefficient on predicted mortality is close to zero (0.004) and not statistically significant. A similar result is obtained in column 2, which restricts attention to low- and middle-income countries only. Columns 3 and 4 repeat these regressions but use the global mortality instrument instead of predicted mortality as the independent variable. Again, there is no significant relationship between the instrument and the share of young people in the population. Moreover, when we look at a longer horizon, as in columns 5-8 with specifications similar to those in columns 1-4 but where $t = 1940$ and 1990, the point estimate on our instrument is actually negative, and again not statistically different from zero. These results thus suggest that, by 1980 and 1990, the main change induced by the international epidemiological was in total population, and not in the demographic structure. Moreover, in Acemoglu and Johnson (2007) it is further shown that the international epidemiological transition affected life expectancy at various ages, not only at birth. This is not surprising, as some of the diseases used in computing predicted mortality –in particular tuberculosis and pneumonia, two of the main killers– affected the entire age distribution.

6.3 2SLS Results

Table 8 presents our main results, which are the 2SLS estimates of the effect of population on conflict. More specifically, our second stage regression is given by equation (4), where population is instrumented by predicted mortality –equation (7)–. As before, we report long-difference regressions for 1940 and 1980 in panel A and panel regressions for $1940 - 1980$ in panel B. This table shows that the effect of population on conflict is positive, and very significant in most specifications. In column 1, the dependent variable is the share of years in internal conflict per decade, as measured by the COW dataset.

The size of the effect (π is estimated to be 0.658) implies that the average change (0.65) in log population leads approximately to 4.3 more years in conflict during the 1980s relative to the 1940s. Compare this to the OLS coefficient of Table 2 (0.303), which implied an effect of around 1.97 more years in conflict in the 1980s compared to the 1940s. For a country like El Salvador, experiencing an increase in population from 1.6 to 4.6 million in this period, the OLS estimate predicts roughly 3.2 more years in conflict while the IV estimate of π implies an effect of roughly 7 more yeas in conflict. We find similar results in the case of the panel regressions for 1940-80 presented in panel B ($\pi = 0.54$ with a p-value of about 0.01).

Columns 2-4 investigate the robustness of this result to using different measures of conflict and the alternative global mortality instrument. In particular, the dependent variables in columns 2 and 3 are the years in internal conflict as a fraction of total years in the reference date as measured by the UCDP/PRIO and Fearon and Laitin datasets, respectively. All the estimated coefficients are positive, and typically significant at less than 1 or 5% according to our cluster-robust wild bootstrap p-values. The more standard p-values from robust standard errors clustered at the country level are somewhat more conservative. Column 4 considers instead the number of years the state failed as a fraction of total years, and column 5 presents regressions for the (log of) battle deaths per year for each reference date. While in the latter the coefficient on population is also positive and significant in both panels, for state failure we find a positive significant effect just in the panel regressions. Finally, columns 6 to 10 repeat the regressions from columns 1-5 but use global mortality as the instrument for population. The results are very similar. This evidence suggests that our results do not depend on the dating of global health interventions.

Table 9 examines the timing of the response of conflict to population growth. In particular, columns 1 to 4 look at different time horizons by estimating long-difference regressions for our baseline measure of conflict on population (instrumented with predicted mortality), where the initial time period is $t = 1940$ and the final date is 1960, 1970, 1980, and 1990. Moreover, column 5 looks at 1970-1997 as the post-period. Consistent with the idea that health improvements and population increase have a lagged effect on social conflict, results are weaker if we only look only at 1940-60 or 1940-70. This result is again not sensitive to the coding of global health interventions, as columns 6-10 reveal.

7 Robustness Checks

7.1 Alternative Samples, Instrument, and World War II

Table 10 presents a number of robustness checks on our results. We focus on our main measure of conflict: share of years in conflict in each reference date according to the COW dataset on civil war. For comparison, column 1 reproduces our main long-difference and panel regression estimates from Table 8. In column 2, we exclude Eastern European Countries. The estimated value of π remains positive, of similar size and statistically significant. Column 3 drops initially rich countries to verify that these results are not driven by the comparison between rich and poor nations and Column 4 uses the global mortality instrument.

Columns 5-7 check whether results are driven by events around World War II. Column 5 excludes the countries demographically most affected by the War, namely Austria, China, Finland, Germany, Italy, Russian Federation (Urlanis, 2003). Column 6 assigns instead the level of conflict of the 1950s to the 1940s. Column 7 simply ignores the war years, and assigns the number of years in conflict from 1946-49 (as a fraction of the 4 years in these interval) to our dependent variable in 1940.

Overall, the coefficient is very stable and retains statistical significance at conventional levels. These robustness checks thus lend more credibility to our baseline estimates.

7.2 Differential Trends

As emphasized before, the main potential threat to our strategy would be that these estimated effects of population on conflict are truly capturing differential trends between countries with different levels of baseline mortality rates. Since these trends may depend on unobservable characteristics, the threat is ultimately untestable. However, we can examine the robustness of our results to the inclusion of differential trends, parametrized as functions of various observable baseline characteristics. If differential trends due to independence, other institutional factors, initial population, and variables commonly emphasized in the empirical studies of civil war do not change our results, we are more confident about our exclusion restriction.

In Table 11, in line with the corresponding first stages in equation (8) and Table 5, our second stage equations take the following form:

$$c_{it} = \pi x_{it} + \zeta_i + \mu_t + \sum_{t=1940}^{1980} \kappa_i' \bar{\omega}_t + \varepsilon_{it}. \tag{9}$$

In column 1, we examine whether the results could be driven by differential trends between countries with "good" and "bad" institutions. While there are many dimensions of institutions, we choose to measure the quality of institutions by average constraints on the executive over 1950-1970. This is a particularly relevant dimension of institutions, since, as noted in Section 2, the commitment problem is a persuasive explanation for civil war. In column 2, κ_i is simply a dummy variable equal to 1 if country i was independent in 1940. Columns 3 and 4 control for differential trends as a function of initial GDP per capita and initial population, respectively.

In columns 5-9, the country characteristics κ_i are variables commonly emphasized as correlates of civil war. A large literature links conflict to natural-resource abundance, in particular oil, gas, and diamonds[12]. A commonly used measure is oil exports divided by GDP or the share of the natural resource sector in GDP (Sachs and Warner, 1995). As Ross (2006) notes, this measure may be a poor proxy of rents in the economy or revenues for the government -key potential drivers of war- since it does not include oil that is produced but consumed domestically, and it does not account for extraction costs which may vary across countries. Also, even at similar levels of production, the numerator tends to be larger in poor countries because poor countries consume less of their own oil. Normalizing by GDP similarly inflates the numbers for poor countries. Motivated by this reasoning, in columns 5-7, κ_i is, respectively: oil production per capita (from Humphreys, 2005), diamond production per capita (from Humphreys, 2005), and oil and gas rents per capita (from Ross, 2006).

A number of theories also suggest that ethnic (or religious) diversity and polarization may be a cause civil war, or at least that they may facilitate surmount the big collective action problems within groups in conflict[13]. Nevertheless, cross-national studies find few

[12] The mechanisms whereby natural resources may lead to conflict are many (for a discussion, see Humphreys, 2005) and disentangling them is not simple. Within-country variation may provide a useful avenue for research, as the study of Dube and Vargas (2008) demonstrates. These authors find that an increase in the international price of coffee reduces violence in coffee-producing regions, while an increase in the international price of petroleum, increases violence in regions with oil reserves and pipelines. This is consistent with the idea that in the capital-intensive sector, an income shock increases the value of controlling the state without increasing wages and the opportunity cost of fighting, while the opposite is true of a shock to the labor-intensive sector (Dal Bó and Dal Bó 2004).

[13] In particular, ethnic identity may generate group cohesion and thus reduce free-riding within armed

differences between the determinants of civil war in general versus "ethnic" civil wars in particular (see Fearon (2006) for a review). This may be surprising, yet it could be driven by the fact that ethnic fragmentation is measured with considerable error. As Blattman and Miguel (2010) point out, the existing proxies may also be theoretically inappropriate and these indices of ethnic fractionalization have been questioned as a meaningful proxy for ethnic tensions (e.g., Posner 2004a, 2004b). Esteban and Ray (1994, 1999) argue that more than fractionalization, a bimodal distribution of preferences or resources—"polarization"—is linked to greater conflict risk. Montalvo and Reynal-Querol (2005) construct measures for polarization and fragmentation and find support for this theory. In columns 8 and 9 we use these measures of ethnic polarization and fractionalization.

Notice that the coefficient remains significant at conventional confidence levels in almost every regression (an exception is column 9 controlling for ethnic fragmentation, where the p-value in the long-difference specification is 12% for the standard cluster-robust variance estimator, but even in this case the bootstrap p-value is around 5% -0.042-, and the panel regressions suggest a significant positive effect). Moreover, the coefficient is quite stable across specifications, ranging from around 0.6 to 0.8 in most long-difference specifications. The sole exception is column 3 that includes a differential trend by initial GDP per capita, yet here the estimated coefficient increases to 1.1. This is reassuring, as we had observed different patterns of conflict for initially rich and poor nations, but this result suggests our estimated impact of population on conflict is unlikely to be explained by differential trends by levels of income. Overall, Table 11 suggests that it is unlikely that the impact of population on conflict from our 2SLS is actually driven by differential trends[14]. The next section investigates a different issue: whether they may be driven by mechanical effects.

groups. As noted in Blattman and Miguel (2010), the motives for such cohesion vary from "primordialist" arguments –stressing the cultural, biological or psychological nature of ethnic cleavages (Horowitz, 1985) or economic models assuming individuals share political preferences or prefer to mingle with co-ethnics (Alesina, Baqir, and Easterly 1999; Alesina and La Ferrara 2000; Esteban and Ray 1999)–; to theories of dense ethnic social networks and low cost information and sanctioning (Caselli and Coleman 2006; Fearon and Laitin 1996; Miguel and Gugerty, 2005); to advantages of ethnic alliances over class alliances in mobilizing for conflict (Esteban and Ray, 2008); and to "modernist" in which ethnic conflict arises when groups excluded from social and political power begin to experience economic modernization (Bates 1986; Ernest Gellner 1983).

[14] Moreover, while Table 11 uses the best available measures of resource abundance and the more theoretically-motivated measures of ethnic diversity, the results do not depend on the exact variable used to measure natural resource abundance or social diversity. This is verified in Appendix Tables 2 and 3, which present the first and second stages, respectively, for specifications similar to those in Table 11 but where alternative measures are used, including: the share of the natural resources in GDP, total (instead of per capita) oil and diamond production, religious polarization and fragmentation, and share of Catholic, Muslim, and Protestant population.

7.3 Further Robustness Checks: Mechanical Effects

Since conventional measures of civil war rely on meeting a battle death threshold, an increase in total population may mechanically increase the number of "detected" civil wars. In this section, we use battle deaths data to verify that this is not driving our results.

Table 12 reports 2SLS regressions of equation (4), where the dependent variable is proportional to the number of battle deaths in country i at time t. In particular, we normalize battle deaths by initial (in 1940) population and take logarithms to calculate c_{it}. For this exercise, we focus on the UCDP/PRIO instead of COW dataset, since the former uses a lower threshold of battle deaths and includes more conflicts. In long-difference regressions (columns 1-3), the estimated coefficient for population is positive. The standard cluster-robust p-value is larger than 10%, but the bootstrap estimates suggest otherwise: the p-value is 4% for the base sample using predicted mortality (column 1), 2.2% when we use global mortality (column 2), and 4.4% when we allow for a differential trend for initially independent countries (column 3). In the panel regressions of columns 4 to 6, the point estimates are similar and we can be more confident about the significance of the effect for the base sample using the predicted (column 4) or global mortality (column 5) instruments. However, in column 6 which includes the interaction with initially independent the coefficient is at best marginally significant at conventional confidence levels.

As an additional check, we consider a simple test to verify whether population changes had an effect on conflict in addition to the mechanical effect that would be expected from size of the population to number of deaths alone. To construct a measure of the deviation of battle deaths from battle deaths predicted by population size, we proceed in two steps. In step 1, we use early data on battle deaths to estimate the cross-sectional relationship between population and battle deaths. In particular, we run a regression of the (log of 1+) battle deaths on log population[15], for an initial baseline reference year. In step 2, we use the estimated coefficients from step 1 and observed population figures for all countries in our sample to construct the "predicted" (log of) battle deaths for each country in each reference year. Our dependent variable is then the deviation of observed (log of) battle deaths from this prediction.

In Panel A of Table 13, we present our standard long-difference and panel regressions where step 1 is calculated using observed battle deaths in 1930. Hence, Panel A relies on data

[15] Analogous to our measures for civil war incidence, the exact form of our dependent variables is $\log\left(1+\frac{battle\ deaths_{it}}{no\ of\ years_{it}}\right)$, where no. of years refers to the number of years with data in the reference year (decade) t for country i.

from COW, as it is our only source with data in 1930. While this source has the disadvantage of using a higher threshold of battle deaths for inclusion as civil war, the advantage is that our baseline regression for the relationship between population and battle deaths is estimated with data that precedes the key period from 1940-1980. In Panel B, however, we also report regressions using data from UCDP/PRIO. In this Panel, step 1 estimates the relationship between battle deaths and population with data from the earliest reference date available, 1940.

Two observations must be made abut this exercise. First, in the regressions to establish the relationship between population size and battle deaths (step 1), we consider only observations with positive battle deaths. Hence this is the relationship between population and fatalities conditional on war. While of course this implies a selection bias, it increases the predicted battle deaths and stacks the cards against finding an effect for the deviation of observed deaths from predicted. Second, while this constitutes an attempt to correct for the mechanical effect of population size, to the extent that IV and OLS estimates of the impact of population on conflict are not too different to each other, our final results for the effect of population on the deviations from predicted should be biased towards zero. The reason is that the cross-sectional relationship even in 1930 or 1940 will include the causal effect of population on Civil War. Imagine, for example, the situation in which OLS and IV estimates are the same. Then one can estimate the same relationship from the initial cross-section and the regression with deviations from predicted values will mechanically give zero. Therefore, to the extent that OLS is similar to IV we will tend to get a downward bias.

With these observations in mind, we take the results from Table 13 to be very suggestive that the results presented thus far are not driven merely by mechanical effects from population size to "detected" wars. Indeed, in Panel A the estimated coefficient for population is significant at either 90% or 95% confidence level for most specifications (depending on whether we follow the standard or bootstrap p-values). In Panel B, the estimated coefficient for population is also positive, but not significant at conventional confidence levels.

Finally, we create new definition of war, based on *relative* threshold of violence. In particular, we define a country-year as in war if its ratio of battle deaths to population is above the 5th, 10th, and 20th percentile among country-years with battle deaths in the sample. With this measure of war incidence, we construct our usual dependent variable as the share of years in war per reference date.

Table 14 reports 2SLS regressions of equation (4) with this measure of c_{it}. The results show positive estimated coefficients on log population, that are somewhat smaller than the corresponding ones using the standard measure of war (for the UCDP/PRIO). For instance, in columns 1 and 2 of Table 14 which use the 5th percentile threshold, the estimated coefficient for (log) population in long differences regressions (Panel A) is 0.492 and 0.527 (when using the predicted and global mortality instruments, respectively). The corresponding coefficients in Panel A of Table 8 (columns 2 and 8), are 0.637 and 0.653. Hence, the impact does appear smaller when looking at the relative threshold of violence to define wars. However, while the coefficients are not significant according to the conventional p-values, the bootstrap p-values still sugest this is a significant positive effect. When increasing the relative threshold, and when looking at panel regressions, the general pattern is that the estimated coefficient on population, while positive, is again smaller, and hence typically not significant, or at best marginally significant at conventional confidence levels.

In sum, while some of the results in this section are less stark than those in the rest of our exercises, the general pattern does not indicate that our results merely mirror a mechanical increase in the number of "detected" civil wars from an increase in total population. In particular, most of our specifications for battle deaths relative to initial population suggest a positive impact of population growth, we present very suggestive evidence that population changes had an effect on conflict in addition to the mechanical effect that would be expected from size of the population to number of deaths alone, and population even has a positive impact in some of our specifications for war defined using relative thresholds of battle deaths to define war.

8 Effects on Regime Changes

The preceding results indicate that large increases in population may have adverse effects on social stability by increasing the incidence of violent conflict. It is also possible to examine if other measures of social change such as political transition are affected by population growth. Indeed, a long tradition emphasizes the role of crises, and social unrest, in inducing political transitions .

To examine this question, Table 15 presents results where the dependent variable is a dummy variable equal to 1 if the country experiences any major regime change over the reference period. The definition of any major regime change comes from PolityIV, and involves significant changes in the polity score either in the direction of more democracy or more autocracy (see Appendix Table 1 for exact definitions). Columns 1-3 vary the sample considered (base sample, excluding Eastern Europe, low and middle income countries only), column 4 uses the global mortality instrument (for the base sample) and columns 5-8 allow for differential trends according to baseline characteristics (institutions, a dummy for independence in 1940, initial GDP per capita, and initial population). As usual, Panel A looks at long difference specifications, and Panel B focuses on panel regressions. The general pattern is very clear: population growth increases the likelihood a major regime changes, and the result is very robust to the various specification checks.

When examining whether these effects are mostly driven by movements towards democracy or autocracy, we find that the most important driving force concerns democratic transitions. Table 16 shows this by presenting the same set of regressions as in Table 15, but where the dependent variable is a dummy variable that takes the value of 1 if the country experiences significant improvement in the democracy score (see Appendix Table 1). Again, the general pattern is that population growth increases the likelihood of significant movements towards democracy.

9 Final remarks

Civil wars create enormous economic costs and human suffering in many countries. This paper shows that the significant (and largely unprecedented) population surges that followed the *international epidemiological transition* that began in the 1940s increased violent social conflict. We interpret our findings as suggestive that larger populations, without a corresponding increase in resources and technology, exacerbate the competition for resources and increase the likelihood of conflict and civil wars in environments where the institutions cannot handle the higher level of disputes. Given the tremendous increase in population in developing countries in the last 100 years, our results highlight an important additional dimension of the implications of population changes in less-developed countries. They may also help explain the puzzling fact that, while the international epidemiological transition produced significant convergence in health conditions around the world, no comparable convergence has been observed in income per capita. Finally, in line with a long tradition emphasizing the role of crises and social unrest in inducing political transitions, we show that the likelihood of major political transitions (and in particular democratic change) also increased as a result of population growth.

10 References

Acemoglu, Daron, Georgy Egorov and Konstantin Sonin (2008), "Dynamics and Stability of Constitutions, Coalitions, and Clubs," NBER Working Paper 14239.

Acemoglu, Daron and James A. Robinson (2001), "A Theory of Political Transitions," *American Economic Review,* 91(4), 938-963.

Acemoglu, Daron, and James A. Robinson (2006), *Economic Origins of Dictatorship and Democracy.* Cambridge and New York: Cambridge University Press.

Acemoglu, Daron and Simon Johnson (2007), "Disease and Development: The Effect of Life Expectancy on Economic Growth," *Journal of Political Economy*, 115, 925-985.

Alesina, Alberto, Reza Baqir, and William Easterly (1999), "Public Goods and Ethnic Divisions", *Quarterly Journal of Economics*, 114(4): 1243–84.

Alesina, Alberto, and Eliana La Ferrara (2000), "Participation in Heterogeneous Communities", *Quarterly Journal of Economics*, 115(3): 847–904.

Bates, Robert H (1986), "Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa." In Doro, Mario E. and Newell M. Schultz. *Governing Black Africa*, New York and London: Africana Publishing Company.

Besley, Timothy J., and Torsten Persson (2008), "The Incidence of Civil War: Theory and Evidence," NBER Working Paper 14585.

Blattman, Christopher and Edward Miguel (2010), "Civil War," *Journal of Economic Literature*, 48:1, 3–57.

Brückner, Markus (2010), "Population Size and Civil Conflict Risk: Is there a Causal Link?," *Economic Journal*, 120(544): 535-550.

Cameron, Colin, Jonah Gelbach and Doug Miller (2008), "Bootstrap-Based Improvements for Inference with Clustered Errors", *Review of Economics and Statistics*, 90(3): 414-427.

Caselli, Francesco, and Wilbur John Coleman (2006), "On the Theory of Ethnic Conflict." National Bureau of Economic Research Working Paper 12125.

Collier, Paul and Anke Hoeffler (1998), "On economic causes of civil war," *Oxford Economic Papers*, 50:4, 563-73.

Collier, Paul and Anke Hoeffler (2004), "Greed and Grievance in Civil War," *Oxford Economic Papers*, 56:4, 563-95.

Diehl, Paul F. and Nils Petter Gleditsch (editors) (2001), *Environmental Conflict*. Boulder,CO: Westview Press.

Dube, Oeindrila, and Juan F. Vargas (2008), "Commodity Price Shocks and Civil Conflict: Evidence from Colombia." Harvard University Center for International Development Graduate Student and Postdoctoral Fellow Working Paper 14.

Easterly, William and Ross Levine (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112(4), 1203-50.

Esteban, Joan, and Debraj Ray (1999), "Conflict and Distribution", *Journal of Economic Theory*, 87(2): 379–415.

Esteban, Joan, and Debraj Ray. 2008. "On the Salience of Ethnic Conflict", *American Economic Review*, 98(5): 2185–2202.

Fearon, James D (1998), "Commitment Problems and the Spread of Ethnic Conflict". In David Lake and Donald Rothchild (eds.) *The International Spread of Ethnic Conflict: Fear, Diffusion, and Escalation*, Princeton: Princeton University Press.

Fearon, James D (2004), "Why Do Some Civil Wars Last So Much Longer Than Others?" *Journal of Peace Research,* 41(3): 275-302.

Fearon, James D (2006), "Ethnic Mobilization and Ethnic Violence." In Weingast, Barry R. and Donald A. Wittman (eds.), *The Oxford Handbook of Political Economy*, 852–68. Oxford and New York: Oxford University Press.

Fearon, James D., and David D. Laitin (1996), "Explaining Interethnic Cooperation." *American Political Science Review*, 90(4): 715–35.

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

Garfinkel, Michelle R. and Stergios Skaperdas (2007), "Economics of Conflict: An Overview," *Handbook of Defense Economics*, 2, 649-709.

Gellner, Ernest (1983), *Nations and Nationalism*. Ithaca and London: Cornell University Press.

Gleditsch, Nils Petter (1998), "Armed Conflict and the Environment: A Critique of the Literature." *Journal of Peace Research,* 35:3, 381-400.

Gleditsch, Nils Petter; Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg & Håvard Strand (2002), "Armed Conflict 1946-2001: A New Dataset," *Journal of Peace Research,* 39(5): 615-637.

Hauge, Wenche and Tanja Ellingsen (1998), "Beyond Environmental Scarcity: Causal Pathways to Conflict," *Journal of Peace Research*, 35(3): 299-317.

Homer-Dixon, Thomas F. (1991), "On the Threshold: Environmental Changes as Causes of Acute Conflict," *International Security*, 16:2, 76 - 116.

Homer-Dixon, Thomas F. (1994), "Environmental Scarcities and Violent Conflict: Evidence from Cases," *International Security*, 19(1): 5-40.

Homer-Dixon, Thomas F. (1999), *Environment, Scarcity, and Violence,* Princeton, N.J.: Princeton University Press.

Horowitz, Donald L. (1985), *Ethnic Groups in Conflict.* Berkeley and London: University of California Press.

Humphreys, Macartan (2005), "Natural Resources, Conflict, and Conflict Resolution: Uncovering the Mechanisms," *Journal of Conflict Resolution*, 49(4): 508-537.

Kaplan, Robert (1994), "The Coming Anarchy," *The Atlantic*, February.

La Porta, Rafael, Lopez de Silanes, Florencio, Shleifer, Andrei and Robert W Vishny (1999), "The Quality of Government," *Journal of Law, Economics, and Organization*, 15(1).

Lacina, Bethany and Nils Petter Gleditsch (2005), "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths", *European Journal of Population,* 21(2–3): 145–166.

MacKay, Angus (1981), "Climate and Popular Unrest in Late Medieval Castile," in T.M. Wigley, M.J. Ingram, and G. Farmer, *Climate and History: Studies in Past Climates and Their Impact on Man*, Cambridge: Cambridge University Press, 356-376.

Maddison, Angus (2006), *The World Economy.* Paris, France: Development Centre of the Organisation for Economic Co-operation and Development

Meredith, Martin (2005), *The Fate of Africa*, New York, N.Y.: Public Affairs.

Miguel, Edward, and Mary Kay Gugerty (2005), "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, 89(11–12): 2325–68.

Miguel, Edward, Shanker Satyanath and Ernest Sergenti (2004), "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 112(4): 725-753

Montalvo, José and Marta Reynal-Querol (2005), "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95(3): 796-816.

Posner, Daniel N (2004a), "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science*, 48(4): 849–63.

Posner, Daniel N (2004b), "The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi." *American Political Science Review*, 98(4): 529–45.

Powell, Robert (2006), "War as a Commitment Problem," *International Organization*, 60(1): 169–203.

Richards, Paul (1996), *Fighting for the Rain Forest: War, Youth and Resources in Sierra Leone*, The International African Institute.

Ross, Michael L (2006), "A Closer Look at Oil, Diamonds, and Civil War", *Annual Review of Political Science*, 9: 265–300.

Ross, Michael L (2008), "Oil, Islam, and Women," *American Political Science Review,* 102(1), 107-123.

Sachs, Jeffrey D. and Andrew M. Warner (1995), "Natural Resource Abundance and Economic Growth," NBER Working Papers 5398.

Sambanis, Nicholas (2002), "A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War," *Defence and Peace Economics*, 13(3): 215-243.

Sambanis, Nicholas (2004), "What Is Civil War? Conceptual and Empirical Complexities of an Operational Definition." *Journal of Conflict Resolution*, 48(6): 814–58.

Sarkees, Meredith R. (2000). "The Correlates of War Data on War: An Update to 1997," *Conflict Management and Peace Science*, 18/1: 123-144

Turner, Michael E. (1984), *Enclosures in Britain, 1750-1830.* London: Macmillan.

Tir, Jaroslav and Paul F. Diehl (1988), "Demographic Pressure and Interstate Conflict: Linking Population Growth and Density to Militarized Disputes and Wars, 1930-89," *Journal of Peace Research*, 35:3, 319-339.

Urlanis, B. (2003) *Wars and Population,* Honolulu, Hawaii: University Press of the Pacific, reprint of 1971 edition.

Wooldridge, Jeffery M (2002), *Econometric Analysis of Cross Section and Panel Data.* Cambridge MA: MIT Press.

Table 1: Descriptive Statistics

	Year	Base Sample	Initially Rich Countries	Initially Middle Income Countries	Initially Poor Countries	Above Median Change in Predicted Mortality 1940 to 1980	Below Median Change in Predicted Mortality 1940 to 1980
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years in conflict/total years, COW	1940	0.051	0.000	0.050	0.088	0.080	0.020
		(0.189)	(0.000)	(0.175)	(0.263)	(0.249)	(0.082)
Years in conflict/total years, COW	1980	0.131	0.000	0.100	0.224	0.252	0.006
		(0.292)	(0.000)	(0.265)	(0.357)	(0.373)	(0.035)
Years in conflict/total years, Uppsala	1940	0.135	0.000	0.115	0.260	0.201	0.067
		(0.309)	(0.000)	(0.276)	(0.417)	(0.369)	(0.219)
Years in conflict/total years, Uppsala	1980	0.277	0.091	0.203	0.444	0.415	0.134
		(0.421)	(0.302)	(0.373)	(0.470)	(0.464)	(0.319)
Years in conflict/total years, Fearon and Laitin	1940	0.121	0.000	0.100	0.238	0.163	0.077
		(0.302)	(0.000)	(0.255)	(0.427)	(0.347)	(0.247)
Years in conflict/total years, Fearon and Laitin	1980	0.242	0.091	0.183	0.376	0.367	0.113
		(0.413)	(0.302)	(0.369)	(0.475)	(0.467)	(0.306)
Years with state failure/total years	1950	0.084	0.000	0.064	0.148	0.116	0.052
		(0.225)	(0.000)	(0.204)	(0.284)	(0.272)	(0.163)
Years with state failure/total years	1980	0.263	0.027	0.176	0.468	0.421	0.100
		(0.404)	(0.090)	(0.352)	(0.457)	(0.454)	(0.265)
Log 1+ Battle Deaths, COW	1940	0.925	0.000	1.138	1.228	1.126	0.708
		(2.696)	(0.000)	(2.759)	(3.425)	(2.829)	(2.585)
Log 1+ Battle Deaths, COW	1980	1.392	0.000	1.028	2.428	2.558	0.191
		(2.856)	(0.000)	(2.644)	(3.381)	(3.513)	(1.080)
Log of population	1940	9.191	9.349	8.871	9.557	9.124	9.261
		(1.499)	(1.344)	(1.349)	(1.756)	(1.617)	(1.393)
Log of population	1980	9.812	9.762	9.393	10.321	9.856	9.768
		(1.384)	(1.293)	(1.238)	(1.461)	(1.484)	(1.294)
Baseline Predicted Mortality	1940	0.469	0.171	0.487	0.626	0.690	0.241
		(0.271)	(0.050)	(0.224)	(0.272)	(0.195)	(0.080)
Global Mortality Index	1940	0.456	0.171	0.482	0.593	0.666	0.238
		(0.258)	(0.050)	(0.222)	(0.252)	(0.184)	(0.079)

NOTE- The table reports the mean values of variables in the samples described in the column heading, with their standard deviations in parentheses. Initially rich countries had log GDP per capita over 8.4 in 1940, middle-income countries had log GDP per capita betweeen 7.37 and 8.4, and low-come countries had log GDP per capita below 7.37 in 1940. Predicted mortality is measured per 100 per year. Columns 6 and 7 report descriptive statistics for subsamples in which change in predicted mortality between 1940 and 1980 was above or below median value in the base sample (-0.405). See the text and Appentix Table 1 for details and definitions.

Table 2. Population and Conflict -OLS Estimates

	Dependent variable is number of years in internal conflict or number of years state failed/total years				Dependent variable:
	(1)	(2)	(3)	(4)	(5)
	COW	Uppsala	Fearon and Laitin	State Failure	Log (1+Battle Deaths/No. of Years)
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1950 and 1980	Just 1940 and 1980
log of population	0.303	0.255	0.236	0.591	2.076
p-value	(0.058)	(0.201)	(0.238)	(0.008)	(0.265)
p-value, wild bootsrap	(0.014)	(0.084)	(0.102)	(0.002)	(0.134)
Observations	102	104	104	125	102
R-squared	0.658	0.656	0.625	0.713	0.561
Number of clusters	50	51	51	62	50
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1950-1980	Panel 1940-1980
log of population	0.265	0.307	0.251	0.528	1.989
p-value	(0.011)	(0.048)	(0.097)	(0.005)	(0.132)
p-value, wild bootsrap	0.000	(0.016)	(0.052)	(0.004)	(0.096)
Observations	307	308	308	256	307
R-squared	0.478	0.629	0.621	0.630	0.496
Number of clusters	63	63	63	63	63

Note.—OLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A is long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B is unbalanced panel with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Table 3. Population and Conflict Controlling for Age Structure -OLS Estimates

	Dependent variable is number of years in internal conflict or number of years state failed/total years								Dependent variable:	
	COW		Uppsala		Fearon and Laitin		State Failure		Log (1+Battle Deaths/No of years)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: long differences	Just 1940 and 1980		Just 1940 and 1980		Just 1940 and 1980		Just 1950 and 1980		Just 1940 and 1980	
share of population 15-34	0.865	-0.670	-0.595	-1.960	-2.091	-3.504	-2.437	-3.560	-4.255	-17.26
p-value	(0.642)	(0.723)	(0.849)	(0.517)	(0.604)	(0.388)	(0.311)	(0.100)	(0.840)	(0.430)
p-value, wild bootsrap	(0.562)	(0.570)	(0.766)	(0.438)	(0.640)	(0.426)	(0.160)	(0.030)	(0.750)	(0.218)
log of population		0.381		0.332		0.344		0.567		3.226
p-value		(0.051)		(0.143)		(0.145)		(0.032)		(0.112)
p-value, wild bootsrap		(0.026)		(0.058)		(0.046)		(0.012)		(0.022)
R-squared	0.618	0.691	0.690	0.724	0.651	0.686	0.629	0.717	0.587	0.639
Observations	86	86	88	88	88	88	90	90	86	86
Number of clusters	43	43	44	44	44	44	45	45	43	43
Panel B: panel regressions	Panel 1940-1980		Panel 1940-1980		Panel 1940-1980		Panel 1950-1980		Panel 1940-1980	
share of population 15-34	-0.465	-1.008	-1.191	-1.716	-1.643	-2.095	-1.538	-2.101	-8.577	-13.51
p-value	(0.437)	(0.108)	(0.247)	(0.096)	(0.306)	(0.198)	(0.269)	(0.100)	(0.193)	(0.061)
p-value, wild bootsrap	(0.346)	(0.074)	(0.286)	(0.108)	(0.384)	(0.216)	(0.270)	(0.076)	(0.146)	(0.042)
log of population		0.330		0.307		0.265		0.504		3.001
p-value		(0.014)		(0.083)		(0.116)		(0.029)		(0.058)
p-value, wild bootsrap		(0.008)		(0.052)		(0.062)		(0.032)		(0.054)
R-squared	0.488	0.538	0.647	0.668	0.669	0.683	0.530	0.575	0.537	0.571
Observations	227	227	228	228	228	228	184	184	227	227
Number of clusters	46	46	46	46	46	46	46	46	46	46

Note.—OLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A is long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B is unbalanced panel with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Table 4. Predicted Mortality and Population: First Stage Estimates and Basic Robustness

	Dependent variable is log population				
	(1)	(2)	(3)	(4)	(5)
	Base Sample	Excluding Eastern Europe	Low and Middle income Countries Only	Global mortality instrument	Base Sample, excluding most affected by WWII
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
baseline predicted mortality	-0.782	-0.700	-0.764	-0.818	-0.811
p-value	(0.000)	(0.001)	(0.008)	(0.000)	(0.000)
p-value, wild bootsrap	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
R-squared	0.988	0.989	0.987	0.988	0.987
Observations	102	92	80	102	94
Number of clusters	50	45	39	50	46
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
baseline predicted mortality	-0.464	-0.402	-0.471	-0.681	-0.476
p-value	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
p-value, wild bootsrap	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
R-squared	0.992	0.993	0.992	0.993	0.991
Number of observations	307	278	252	307	279
Number of clusters	63	57	52	63	57

Note.—OLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A is long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B is unbalanced panel with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Table 5. Predicted Mortality and Population: First Stage Estimates and Robustness to Differential Trends

	Dependent variable is log of population								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Interaction of post-year dummies with…								
	Institutions	Independent in 1940	Log GDP per capita in 1930	Log Population in 1930	Oil Production per Capita in 1960	Diamond Production per Capita in 1960	Oil and Gas rents per Capita in 1960	Ethnic Polarization	Ethnic Fragmentation
Panel A: Long Differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
Baseline predicted mortality	-0.730	-0.815	-0.600	-0.762	-0.782	-0.776	-0.808	-0.599	-0.587
p-value	(0.004)	(0.002)	(0.061)	(0.000)	(0.000)	(0.000)	(0.000)	(0.007)	(0.018)
p-value, wild bootsrap	(0.002)	(0.002)	(0.012)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
p-value for post year dummy x …	0.605	0.700	0.240	0.0659	9.64e-10	1.77e-07	1.50e-07	0.0413	0.185
R-squared	0.988	0.988	0.989	0.989	0.989	0.989	0.989	0.991	0.990
Observations	102	102	100	102	102	102	102	96	96
Number of clusters	50	50	49	50	50	50	50	47	47
Panel B: Panel	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
Baseline predicted mortality	-0.436	-0.484	-0.361	-0.446	-0.463	-0.460	-0.478	-0.337	-0.330
p-value	(0.001)	(0.001)	(0.033)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.008)
p-value, wild bootsrap	(0.002)	(0.002)	(0.028)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.008)
p-value for post year dummies x …	0.000358	0.0155	1.90e-08	0.111	0	2.40e-10	4.22e-10	0.000893	0.000488
R-squared	0.993	0.992	0.994	0.992	0.993	0.993	0.993	0.994	0.994
Observations	307	307	267	265	307	307	307	280	280
Number of clusters	63	63	53	52	63	63	63	57	57

Note.—OLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A is long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B is unbalanced panel with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Table 6. Reduced Forms and Falsification Exercises

	Base Sample	Low and Middle Income Countries Only	Base Sample	Low and Middle Income Countries Only	Base Sample	Low and Middle Income Countries Only
	(1)	(2)	(3)	(4)	(5)	(6)
	Reduced Form		Falsification Exercises			
Dependent variable is: --	Change in years in conflict/total years from 1940 to 1980		Change in years in conflict/total years from 1900 to 1940		Change in Log Population from 1900 to 1940	
Change in Predicted Mortality from 1940 to 1980	-0.516	-0.643	0.0592	0.122	-0.189	-0.198
p-value	(0.002)	(0.003)	(0.156)	(0.328)	(0.176)	(0.318)
p-value, wild bootsrap	(0.006)	(0.006)	(0.110)	(0.388)	(0.226)	(0.352)
Observations	52	41	36	28	52	41
R-squared	0.203	0.215	0.007	0.019	0.033	0.029

Note.—OLS regressions. Regular p-values for t-statistics with robust standard errors are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). See the text and Appendix Table 1 for definitions and details.

Table 7. Predicted Mortality and Age Structure

	Dependent variable is Share of Poulation of ages 15-34							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Base Sample	Low and Middle income Countries Only	Base Sample	Low and Middle income Countries Only	Base Sample	Low and Middle income Countries Only	Base Sample	Low and Middle income Countries Only
long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1990	Just 1940 and 1990	Just 1940 and 1990	Just 1940 and 1990
baseline predicted mortality	0.00384	-0.000329			-0.0285	-0.0299		
p-value	(0.871)	(0.993)			(0.300)	(0.550)		
p-value, wild bootsrap	(0.808)	(0.978)			(0.144)	(0.510)		
global mortality instrument			0.00462	0.000541			-0.0300	-0.0317
p-value			(0.859)	(0.990)			(0.314)	(0.568)
p-value, wild bootsrap			(0.798)	(0.962)			(0.194)	(0.532)
R-squared	0.587	0.533	0.588	0.533	0.571	0.526	0.571	0.526
Observations	86	64	86	64	86	64	86	64
Number of clusters	43	32	43	32	43	32	43	32

Note.—OLS regressions. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). See the text and Appendix Table 1 for definitions and details.

Table 8. The Effect of Population and Conflict - 2SLS Estimates

	Dependent variable is years in internal conflict or years state failed/total years (cols 1-4, 5-8) and Log of 1+Battle Deaths/No. of years (cols 5 and 10)									
	Baseline Predicted Mortality					Using Global Mortality Rate				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	COW	Uppsala	Fearon and Laitin	State Failure	Log (1+Battle Deaths/Total Years)	COW	Uppsala	Fearon and Laitin	State Failure	Log (1+Battle Deaths/Total Years)
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1950 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1950 and 1980	Just 1940 and 1980
Log of population	0.658	0.637	0.879	0.0315	6.774	0.649	0.653	0.872	0.681	6.646
p-value	(0.034)	(0.089)	(0.051)	(0.985)	(0.040)	(0.039)	(0.093)	(0.053)	(0.151)	(0.044)
p-value, wild bootsrap	(0.006)	(0.020)	0.000	(0.986)	(0.008)	(0.008)	(0.010)	(0.010)	(0.060)	(0.008)
Observations	102	104	104	125	102	102	104	104	125	102
Number of clusters	50	51	51	62	50	50	51	51	62	50
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1950-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1950-1980	Panel 1940-1980
Log of population	0.540	0.425	0.761	0.737	6.229	0.540	0.425	0.761	0.737	6.229
p-value	(0.013)	(0.141)	(0.051)	(0.053)	(0.024)	(0.013)	(0.141)	(0.051)	(0.053)	(0.024)
p-value, wild bootsrap	(0.014)	(0.092)	(0.020)	(0.046)	(0.018)	(0.006)	(0.088)	(0.022)	(0.022)	(0.020)
Observations	307	308	308	256	307	307	308	308	256	307
Number of clusters	63	63	63	63	63	63	63	63	63	63

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A is long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B is unbalanced panel with one observation per decade. First stages, for the sample with data on years in conflict according to COW, in column 1 and column 4 of Table 4. See the text and Appendix Table 1 for definitions and details.

Table 9. Timing the Effect of Population on Conflict -2SLS Estimates

	Dependent variable is years in conflict/total years according to Correlates of War -COW-									
	Baseline Predicted Mortality					Using Global Mortality Rate				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Just 1940 and 1960	Just 1940 and 1970	Just 1940 and 1980	Just 1940 and 1990	Just 1940s and 1970-1997	Just 1940 and 1960	Just 1940 and 1970	Just 1940 and 1980	Just 1940 and 1990	Just 1940s and 1970-1997
log of population	0.111	0.471	0.658	0.464	0.749	0.114	0.456	0.649	0.436	0.724
p-value	(0.630)	(0.105)	(0.034)	(0.073)	(0.030)	(0.668)	(0.110)	(0.039)	(0.063)	(0.029)
p-value, wild bootsrap	(0.490)	(0.012)	(0.004)	(0.022)	(0.572)	(0.022)	(0.010)	(0.018)	(0.004)	(0.012)
Observations	102	102	102	102	102	102	102	102	102	102
Number of clusters	50	50	50	50	50	50	50	50	50	50

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Long-difference specifications with two observations per country, one for the intial date and one for the final date. First stage for columns 1 to 5 in column 1 of Table 4-Panel A, and for columns 6-10 in column 4 of Table 4-Panel A. for See the text and Appendix Table 1 for definitions and details.

Table 10. The Effect of Population on Conflict: Basic Robustness Checks -2SLS Estimates

	Dependent variable is years in conflict/total years according to Correlates of War -COW-						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Base Sample	Excluding Eastern Europe	Low and Middle income Countries Only	Global mortality instrument	Base Sample, excluding most affected by WWII	Base Sample, assign 1950 to 1940	Base Sample, assign 1946-1949 to 1940
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
log of population	0.658	0.739	0.841	0.649	0.650	0.524	0.687
p-value	(0.034)	(0.041)	(0.056)	(0.039)	(0.038)	(0.098)	(0.040)
p-value, wild bootsrap	(0.010)	(0.006)	(0.008)	(0.006)	(0.004)	(0.042)	(0.004)
Observations	102	92	80	102	94	102	102
Number of clusters	50	45	39	50	46	50	50
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
log of population	0.509	0.598	0.683	0.540	0.507	0.282	0.557
p-value	(0.011)	(0.015)	(0.022)	(0.013)	(0.013)	(0.120)	(0.021)
p-value, wild bootsrap	(0.006)	(0.006)	(0.008)	(0.014)	(0.008)	(0.076)	(0.004)
Observations	307	278	252	307	279	307	307
Number of clusters	63	57	52	63	57	63	63

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A is long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B is unbalanced panel with one observation per decade. First stages for columns 1-5 in columns 1-5 of Table 4. First stages of columns 6-7 are in column 1 of Table 4. See the text and Appendix Table 1 for definitions and details.

Table 11. The Effect of Population on Conflict: Differential Trends -2SLS Estimates

	Dependent variable is years in conflict/total years according to COW								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Interaction of post-year dummies with…								
	Institutions	Independent in 1940	Log GDP per capita in 1930	Log Population in 1930	Oil Production per Capita in 1960	Diamond Production per Capita in 1960	Oil and Gas rents per Capita in 1960	Ethnic Polarization	Ethnic Fragmentation
Panel A: Long Differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
Log of population	0.791	0.589	1.127	0.665	0.658	0.665	0.635	0.805	0.748
p-value	(0.037)	(0.069)	(0.095)	(0.034)	(0.033)	(0.033)	(0.033)	(0.064)	(0.124)
p-value, wild bootsrap	(0.008)	(0.012)	(0.008)	(0.002)	(0.008)	(0.012)	(0.008)	(0.036)	(0.042)
p-value for post year dummy x	0.314	0.527	0.260	0.663	0.0337	0.0393	0.0685	0.467	0.819
Observations	102	102	100	102	102	102	102	96	96
Number of clusters	50	50	49	50	50	50	50	47	47
Panel B: Panel	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
Log of population	0.545	0.454	0.930	0.517	0.510	0.513	0.494	0.599	0.503
p-value	(0.029)	(0.011)	(0.038)	(0.011)	(0.010)	(0.010)	(0.010)	(0.038)	(0.077)
p-value, wild bootsrap	(0.012)	(0.004)	(0.002)	(0.004)	(0.006)	(0.010)	(0.004)	(0.026)	(0.044)
p-value for post year dummies x …	0.424	0.722	0.0176	0.249	0.157	0.0748	0.231	0.647	0.168
Observations	307	307	267	265	307	307	307	280	280
Number of clusters	63	63	53	52	63	63	63	57	57

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A is long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B is unbalanced panel with one observation per decade. First stages in Table 5. See the text and Appendix Table 1 for definitions and details.

Table 12. The Effect of Population on Conflict: Mechanical Effects I -2SLS Estimates

	Dependent variable is Log (1+Battle Deaths/Population in 1940), Uppsala					
	long differences			panel regressions		
	Base Sample	Global Mortality Instrument	Interaction with Independent in 1940	Base Sample	Global Mortality Instrument	Interaction with Independent in 1940
	(1)	(2)	(3)	(4)	(5)	(6)
	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
log of population	1.363	1.407	1.462	1.088	1.275	1.054
p-value	(0.134)	(0.136)	(0.179)	(0.057)	(0.034)	(0.154)
p-value, wild bootsrap	(0.040)	(0.022)	(0.044)	(0.040)	(0.038)	(0.114)
p-value for Year dummies x Independent in 1940			0.739			0.640
Observations	104	104	104	273	273	273
Number of clusters	51	51	51	54	54	54

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Columns 1-3 are long-difference specifications with two observations per country, one for the intial date and one for the final date. Columns 4-6 are unbalanced panel with one observation per decade. First stage for columns 1 and 4 in column 1 of Table 4, for columns 2 and 5 in column 4 of Table 4, and for columns 3 and 6 in column 2 of Table 5. See the text and Appendix Table 1 for definitions and details.

Table 13. The Effect of Population on Conflict: Mechanical Effects II -2SLS Estimates

	Base Sample	Global Mortality Instrument	Interaction with Independent in 1940	Base Sample	Global Mortality Instrument	Interaction with Independent in 1940
	Dependent variable is Deviation of Observed Log (1+ Battle Deaths/No. of years) from Predicted, Uppsala					
	long differences			panel regressions		
	(1)	(2)	(3)	(4)	(5)	(6)
	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
Panel A: Predicted based on COW, 1930						
log of population	5.815	5.687	4.875	6.199	5.270	4.612
p-value	(0.077)	(0.085)	(0.152)	(0.030)	(0.055)	(0.085)
p-value, wild bootsrap	(0.020)	(0.034)	(0.046)	(0.018)	(0.040)	(0.036)
p-value for Year dummies x Independent in 1940			0.357			0.101
Observations	103	103	103	307	307	307
Number of clusters	51	51	51	63	63	63
Panel B: Predicted based on Upssala, 1940						
log of population	3.560	3.831	4.119	2.830	2.880	3.065
p-value	(0.318)	(0.297)	(0.342)	(0.421)	(0.392)	(0.455)
p-value, wild bootsrap	(0.160)	(0.150)	(0.192)	(0.374)	(0.338)	(0.414)
p-value for Year dummies x Independent in 1940			0.738			0.367
Observations	104	104	104	308	308	308
Number of clusters	51	51	51	63	63	63

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). To calculate the dependent variable, we: (i) run a cross-sectional regression of the log of (1+ battle deaths per year/no. of years) on log population for countries with positive battle deaths in 1930 (Panel A) or 1940 (Panel B); (ii) next, we use the estimated coefficients and observed population figures to estimate a predicted value of our dependent variable in each reference year; (iii) finally, we calculate the deviation of the observed vale of log of (1+ battle deaths per year/no.of years) from this prediction and use this as the dependent variable in the reported regressions. In Panel A, step (i) uses COW as the source for the dependent variable, and step (iii) constructs the deviation relative to observed battle deaths from COW. In Panel B, battle deaths information is from Uppsala. Columns 1-3 are long-difference specifications with two observations per country, one for the intial date and one for the final date. Columns 4-6 are unbalanced panels with one observation per decade. First stage for columns 1 and 4 in column 1 of Table 4, for columns 2 and 5 in column 4 of Table 4, and for columns 3 and 6 in column 2 of Table 5. See the text and Appendix Table 1 for definitions and details.

Table 14. The Effect of Population on Conflict: Mechanical Effects III -2SLS Estimates

	Dependent variable is Years in internal conflict/total years, Uppsala					
	War if Above 5th Percentile of Battle Deaths/Population		War if Above 10th Percentile of Battle Deaths/Population		War if Above 20th Percentile of Battle Deaths/Population	
	Predicted Mortality	Global Mortality	Predicted Mortality	Global Mortality	Predicted Mortality	Global Mortality
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
log of population	0.492	0.527	0.407	0.440	0.383	0.427
p-value	(0.186)	(0.164)	(0.245)	(0.212)	(0.322)	(0.256)
p-value, wild bootsrap	(0.070)	(0.052)	(0.144)	(0.082)	(0.192)	(0.138)
Observations	104	104	104	104	104	104
Number of clusters	51	51	51	51	51	51
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
log of population	0.215	0.299	0.144	0.252	0.119	0.253
p-value	(0.505)	(0.310)	(0.653)	(0.360)	(0.726)	(0.367)
p-value, wild bootsrap	(0.458)	(0.274)	(0.570)	(0.300)	(0.744)	(0.306)
Observations	308	308	308	308	308	308
Number of clusters	63	63	63	63	63	63

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A are long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B are unbalanced panels with one observation per decade. First stage for odd columns in column 1 of Table 4, and for even columns in column 4 of Table 4. for See the text and Appendix Table 1 for definitions and details.

Table 15. The Effect of Population on Regime Changes -2SLS Estimates

	Dependent variable is Dummy=1 if any major regime change in reference period							
	Base Sample	Excluding Eastern Europe	Low and Middle income Countries Only	Global mortality instrument	Interaction with Institutions	Interaction with Independent in 1940	Interaction with (log) GDP per capita in 1930	Interaction with Log Population in 1930
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
log of population	0.576	0.471	0.963	0.557	0.892	0.495	0.777	0.594
p-value	(0.145)	(0.262)	(0.094)	(0.159)	(0.056)	(0.206)	(0.278)	(0.122)
p-value, wild bootsrap	(0.048)	(0.140)	(0.004)	(0.052)	(0.010)	(0.076)	(0.094)	(0.022)
p-value for post year dummy x …								
Institutions					0.133			
Independent in 1940						0.586		
Initial GDP							0.639	
Continent Dummies								0.378
Observations	104	94	82	104	104	104	102	104
Number of clusters	51	46	40	51	51	51	50	51
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
log of population	1.100	0.968	1.532	0.892	1.458	0.923	1.402	1.149
p-value	(0.008)	(0.030)	(0.012)	(0.015)	(0.009)	(0.020)	(0.110)	(0.008)
p-value, wild bootsrap	(0.004)	(0.014)	(0.006)	(0.006)	0.000	(0.004)	(0.032)	0.000
p-value for year dummies x …								
Institutions					0.484			
Independent in 1940						0.380		
Initial GDP							0.184	
Continent Dummies								0.594
Observations	308	279	253	308	308	308	268	266
Number of clusters	63	57	52	63	63	63	53	52

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A are long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B are unbalanced panels with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Table 16. The Effect of Population on Democratic Change -2SLS Estimates

	Dependent variable is Dummy=1 if Major democractic change in reference period							
	Base Sample	Excluding Eastern Europe	Low and Middle income Countries Only	Global mortality instrument	Interaction with Institutions	Interaction with Independent in 1940	Interaction with (log) GDP per capita in 1930	Interaction with Log Population in 1930
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
log of population	0.457	0.391	0.636	0.439	0.610	0.421	0.768	0.474
p-value	(0.097)	(0.168)	(0.115)	(0.102)	(0.088)	(0.112)	(0.202)	(0.064)
p-value, wild bootsrap	(0.030)	(0.082)	(0.028)	(0.020)	(0.010)	(0.036)	(0.048)	(0.004)
p-value for post year dummy x …								
Institutions					0.368			
Independent in 1940						0.803		
Initial GDP							0.438	
Continent Dummies								0.337
Observations	104	94	82	104	104	104	102	104
Number of clusters	51	46	40	51	51	51	50	51
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
log of population	0.663	0.561	0.928	0.526	0.898	0.755	1.185	0.701
p-value	(0.008)	(0.021)	(0.030)	(0.014)	(0.031)	(0.010)	(0.076)	(0.015)
p-value, wild bootsrap	0.000	(0.002)	0.000	(0.002)	(0.008)	(0.010)	(0.018)	(0.008)
p-value for year dummies x …								
Institutions					0.114			
Independent in 1940						0.489		
Initial GDP							0.583	
Continent Dummies								0.518
Observations	308	279	253	308	308	308	268	266
Number of clusters	63	57	52	63	63	63	53	52

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A are long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B are unbalanced panels with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Appendix Table 1. Variable Sources and Description

Variable	Description	Source
	Social Conflict	
Years in conflict/Total years	Ratio of the number of years with an internal conflict to total years assigned to reference date. Assignment of years to reference dates and exact definition of internal conflict varies by data source, as detailed below.	
COW	Number of years with internal conflict defined as combat between a State and other actor, with three categories included: civil war for control of central government, conflict over local issues, and inter-communal conflict. The threshold for inclusion is 1,000 battle-related deaths, civilian or military, but with massacres excluded. Assignment to reference dates: 1900=1900-09, 1940=1940-49, 1950=1950-59, ..., 1980=1980-89, 1990=90-1999.	Correlates of War (COW) Intra-State wars dataset. See Sarkees (2000).
Uppsala	Number of years with any incidence of an "internal armed conflict" or of "internationalized internal armed conflict". Armed conflict is defined to include all contested incompatibilities that concern government or territory or both where the use of armed force between two parties results in at least 25 battle-related deaths. Of the two parties, at least one is the government of a state. Assignment to reference dates: 1940=1946-49, 1950=1950-59, …, 1980=1980-89, 1990=1990-99.	Backdating of the Uppsala Conflict Data Project (UCDP) database made in conjunction with the International Peace Research Institute, Oslo (PRIO). See Geldtisch et. al (2002).
Fearon and Laitin	Number of years with violent civil conflicts that: (1) involved fighting between agents of (or claimants to) a state and organized, nonstate groups who sought either to take control of a government, to take power in a region, or to use violence to change government policies, (2) killed at least 1,000 over its course, with a yearly average of at least 100, (3) At least 100 were killed on both sides (including civilians attacked by rebels). Counts anticolonial wars as occurring within the empire in question (e.g., Algeria is assigned to France). Assignment to reerence dates: : 1940=1945-49, 1950=1950-59, …, 1980=1980-89, 1990=1990-99.	Fearon and Laitin (2003)
State Failure	Number of years where state failure dummy equals 1. Coded for all countries with populations greater than 500,000 from 1955 to 2006, and equals 1 if there is any serious instability in a given year. Four types of political instability are included: ethnic wars, revolutionary wars, genocides and politicides, and adverse regime changes. Assignment to reference dates: 1950=1955-59, 1960=1960-69,…, 1980=1980-89, 1990=1990-99.	State Failure Task Force, phase 3 replication dataset.
Log(1+Battle Deaths)	"Best estimate" of annual battle-related deaths for use with the COW dataset (version 2) and UCDP/PRIO dataset (version 3). Assignment to reference dates as in the respective conflict dataset.	Center for the Study of Civil War --CSCW-Battle Deaths Dataset. See Lacina and Gleditsch (2005).
Regime change	Dummy variable, takes value of 1 if the country experiences a regime change in the reference period. We consider in particular: a dummy for Major Democratic Transition, defined by Polity IV as six points or greater increase in the POLITY score over a period of three years or less including a shift from an autocratic POLITY value (-10 to 0) to a partial democratic POLITY value (+1 to +6) or full democratic POLITY value (+7 to +10) or a shift from a partial democratic value to a full democratic value; a major adverse regime change, defined by PolityIV as a six or more point decrease in POLITY score or an interregnal period denoting a collapse of central state authority or a revolutionary transformation in the mode of governance (not a democratic transition); and finally a dummy for any major rgime change, which equals one for either major democratic or adverse transitions.	Polity IV.
	Population	
Log of Population	Total Population per country in 1900, 1940, 1950, 1960, 1970, 1980, 1990, 1990.	Maddison (2003)
Share of population 15-34, Share of other age groups	Percentage of the population in each age group for 1940, 1950, 1960, 1970, and 1980. Other age groups included are 0-4, 5-9, 10-14,…, 75-79, and over 80.	From 1950 onwards, UN demographic database (http://esa.un.org/unpp). For 1940, UN Demographic Yearbook 1948 (United Nations 1949, Table 4, pp. 108-158). We use data for 1940 or the closest available year or range of years.

Appendix Table 1 (continued). Variable Sources and Description

Variable	Description	Source
Health		
Predicted Mortality Instrument	Sum of country's initial (in 1940) mortality rate from 15 diseases until there is a global intervention, and after the global intervention, the mortality rate from the disease in question declines to the frontier mortality rate. See paper for mathematical formula. 15 diseases are (in rough descending order of importance): malaria, pneumonia, and tuberculosis; influenza, cholera, typhoid, smallpox, shigella dysentery, whooping cough, measles (rubeola), dyphteria, scarlet fever, yellow fever, plague, typhus.	Acemoglu and Johnson (2007), based on various sources (Summary of International Vital Statistics, Federal Security Agency (1947) of the US government, League of Nations sources, World Health Organization (1951), UN Demographic Yearbooks.)
Global Mortality Instrument	Sum of the products of each country's initial (in 1940) mortality rate from 10 diseases and the ratio between the global mortality at time t to the initial (in 1940) global mortality from the disease in question. See paper for mathematical formula. Like in Acemoglu and Johnson (2007), diseases are as for Predicted Mortality except yellow fever and dysentery/diarrhea for which it was not possible to track the diseases through changes in the classification of death over time. We also exclude cholera, typhoid, and plague since their were often not available for our extended sample of countries. Global Mortality is the unweighted average accross countries in the sample of 59 countries (47 non-Eastern European countries and 12 countries with life expectancy data since 1950, see "Base Sample" below) in Acemoglu and Johnson (2007) for which we have an evolution of mortality rates through time.	Acemoglu and Johnson (2007)
Others		
Base Sample	Our sample includes Acemoglu and Johnson's (2007) list of 47 non-Eastern Europe countries (Argentina, Australia, Austria, Bangladesh, Belgium, Brazil, Canada, Chile, China, Colombia, Costa Rica, Denmark, Ecuador, El Salvador, Finland, France, Germany, Greece, Guatemala, Honduras, India, Indonesia, Ireland, Italy, South Korea, Rep., Malaysia, Mexico, Myanmar, Netherlands, New Zealand, Nicaragua, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Portugal, Spain, Sri Lanka, Sweden, Switzerland, Thailand, United Kingdom, United States, Uruguay, Venezuela); the set of 12 additional countries for which they have life expectancy data since 1950 (Algeria, Bolivia, Egypt, Iran, Iraq, Lebanon, Morocco, Singapore, South Africa, Tunisia, Turkey and Vietnam); and 6 countries from Eastern Europe (Bulgaria, Czech Republic, Hungary, Poland, Romania, and Russian Federation). This implies a total of 65 countries, but not all have all variables for all years. In particular, we lack 1940 population data for Algeria, Egypt, Iran, Iraq, Lebanon, Morocco, Russia, Singapore, South Africa, Tunisia, and Vietnam. Also, Austria is excluded in 1940 when the dependent variables are from COW since it enters the COW state system in the 1950s.	
Initially rich, middle-income, and poor countries	Each category is defined using the top, middle, and lowest third group of countries in the base sample based on income in 1940. Initially rich countries had log GDP per capita over 8.4. in 1940; middle income had log GDP per capita betweeen 7.37 and 8.4; and low income countries had log GDP per capita below 7.37 in 1940.	
Country clusters	For clustered standard errors, Bangladesh, India and Pakistan are considered a single cluster.	
Countries most affected by World War II	Austria, China, Finland, Germany, Italy, Russian Federation	Acemoglu and Johnson (2007)
Controls and Baseline characteristics		
Institutions	Average of constraints on the executive in 1950, 1960 and 1970	Polity IV
Independent in 1940	=1 if country is independent in 1940, 0 otherwise	Own coding.
Initial GDP	Logarithm of GDP per capita in 1930	Maddison (2003)
Countries most affected by World War II	Austria, China, Finland, Germany, Italy, Russian Federation	Acemoglu and Johnson (2007) based on Urlanis (2003)
Natural Resources	Oil Production and Oil Production per capita in 1960 Diamond Production and Diamond Production per capita in 1960 Share of natural resource sector in GNP in 1970 Share of mineral produciton in GNP in 1971 Oil and gas rents per capita in 1960	Humphreys (2005) Humphreys (2005) Sachs and Warner (1995) Sachs and Warner (1995) Ross (2008) (Michael L. Ross, 2008, "Replication data for: Oil, Islam, and Women ", http://hdl.handle.net/1902.1/14307 UNF:5:fsZ56s2dvxP26at+iCdOhg== V1)
Ethnolinguistic fractionalization and polarization and religious composition	Ethnic Polarization Ethnic Fragmentation Religious Polarization Religious Fragmentation Ethnolinguistic fractionalization index (from 0 to 1). Average value of five indices based on ethnic or linguistic characteristics of the population. Share of Muslim, Catholic and Protestant Populations in 1980.	Montalvo and Reynal-Querol (2005) Montalvo and Reynal-Querol (2005) Montalvo and Reynal-Querol (2005) Montalvo and Reynal-Querol (2005) Easterly and Levine (1997) La Porta et al (1999)

Appendix Table 2. Predicted Mortality and Population: First Stage Estimates and Additional Robustness to Differential Trends

	Dependent variable is log of population							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Interaction of post-year dummies with…							
	Share of GDP in Natural Resource Sector in 1970	Share of Mineral Production in GNP in 1971	Oil Production in 1960	Diamond Production in 1960	Religious Polarization	Religious Fragmentation	Average Ethnolinguistic Fragmentation	Share of Catholic, Muslim, Protestant in 1980
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
Baseline Predicted Mortality	-0.495	-0.661	-0.796	-0.772	-0.562	-0.636	-0.777	-0.584
p-value	(0.081)	(0.003)	(0.000)	(0.000)	(0.029)	(0.005)	(0.009)	(0.034)
p-value, wild bootsrap	(0.016)	(0.002)	(0.002)	(0.002)	(0.012)	(0.002)	(0.002)	(0.006)
p-value for post year dummy x …	0.967	0.331	4.24e-06	0.0759	0.169	0.182	0.973	0.465
R-squared	0.992	0.990	0.989	0.989	0.990	0.989	0.988	0.989
Observations	68	92	102	102	96	96	102	100
Number of clusters	33	45	50	50	47	47	50	49
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
Baseline Predicted Mortality	-0.268	-0.378	-0.472	-0.458	-0.321	-0.366	-0.429	-0.357
p-value	(0.047)	(0.001)	(0.000)	(0.000)	(0.015)	(0.002)	(0.002)	(0.012)
p-value, wild bootsrap	(0.026)	(0.002)	(0.002)	(0.002)	(0.020)	(0.006)	(0.002)	(0.014)
p-value for post year dummies x …	1.69e-05	0.000486	3.10e-10	3.07e-05	5.38e-06	0.00135	0.00339	1.55e-08
R-Squared	0.995	0.993	0.993	0.992	0.994	0.993	0.992	0.994
Observations	201	277	307	307	280	280	302	309
Number of clusters	39	55	63	63	57	57	60	62

Note.—OLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A are long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B are unbalanced panels with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Appendix Table 3. The Effect of Population on Conflict: Additional Differential Trends -2SLS Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable is years in conflict/total years according to COW							
	Interaction of post-year dummies with…							
	Share of GDP in Natural Resource Sector in 1970	Share of Mineral Production in GNP in 1971	Oil Production in 1960	Diamond Production in 1960	Religious Polarization	Religious Fragmentation	Average Ethnolinguistic Fragmentation	Share of Catholic, Muslim, Protestant in 1980
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
log of population	1.526	0.811	0.645	0.669	0.920	0.834	0.608	1.054
p-value	(0.101)	(0.041)	(0.031)	(0.032)	(0.111)	(0.058)	(0.135)	(0.103)
p-value, wild bootsrap	(0.002)	(0.006)	(0.010)	(0.008)	(0.042)	(0.024)	(0.046)	(0.010)
p-value for post year dummy x …	0.248	0.165	0.0401	0.124	0.491	0.383	0.810	0.444
Observations	68	92	102	102	96	96	102	100
Number of clusters	33	45	50	50	47	47	50	49
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
log of population	1.516	0.627	0.501	0.517	0.650	0.614	0.369	0.730
p-value	(0.063)	(0.023)	(0.010)	(0.010)	(0.088)	(0.035)	(0.062)	(0.078)
p-value, wild bootsrap	(0.014)	(0.020)	(0.002)	(0.008)	(0.054)	(0.014)	(0.056)	(0.024)
p-value for post year dummies x …	0.349	0.0854	0.208	0.0745	0.637	0.469	0.241	0.817
Observations	194	270	307	307	280	280	295	302
Number of clusters	39	55	63	63	57	57	60	62

Note.—2SLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A are long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B are unbalanced panels with one observation per decade. First stages in Appendix Table 2. See the text and Appendix Table 1 for definitions and details.

Appendix Table 4. Reduced Forms Robustness I: Population and Predicted Mortality

	Dependent variable is years in conflict/total years according to COW						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Base Sample	Excluding Eastern Europe	Low and Middle income Countries Only	Global mortality instrument	Base Sample, excluding most affected by WWII	Base Sample, assign 1950 to 1940	Base Sample, assign 1946-1949 to 1940
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
baseline predicted mortality	-0.515	-0.517	-0.642	-0.531	-0.527	-0.410	-0.537
p-value	(0.028)	(0.029)	(0.033)	(0.035)	(0.031)	(0.098)	(0.030)
p-value, wild bootsrap	(0.006)	(0.006)	(0.006)	(0.016)	(0.008)	(0.034)	(0.006)
R-squared	0.683	0.679	0.681	0.680	0.693	0.683	0.655
Observations	102	92	80	102	94	102	102
Number of clusters	50	45	39	50	46	50	50
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
baseline predicted mortality	-0.245	-0.247	-0.306	-0.367	-0.250	-0.137	-0.265
p-value	(0.003)	(0.003)	(0.004)	(0.012)	(0.003)	(0.098)	(0.005)
p-value, wild bootsrap	(0.004)	(0.002)	(0.002)	(0.008)	(0.002)	(0.068)	(0.002)
R-squared	0.451	0.447	0.445	0.475	0.456	0.482	0.445
Observations	313	283	258	308	284	313	313
Number of clusters	63	57	52	63	57	63	63

Note.—OLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A are long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B are unbalanced panels with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Appendix Table 5. Reduced Forms Robustness II: Population and Predicted Mortality

	Dependent variable is years in conflict/total years according to COW								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Interaction of post-year dummies with…								
	Institutions	Independent in 1940	Log GDP per capita in 1930	Log Population in 1930	Oil Production per Capita in 1960	Diamond Production per Capita in 1960	Oil and Gas rents per Capita in 1960	Ethnic Polarization	Ethnic Fragmentation
Panel A: Long Differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
Baseline predicted mortality	-0.578	-0.480	-0.676	-0.507	-0.515	-0.516	-0.513	-0.482	-0.439
p-value	(0.023)	(0.050)	(0.044)	(0.029)	(0.030)	(0.029)	(0.030)	(0.061)	(0.109)
p-value, wild bootsrap	(0.006)	(0.018)	(0.028)	(0.008)	(0.004)	(0.008)	(0.008)	(0.004)	(0.028)
p-value for post year dummy x …				0.442	0.277	0.164	0.743	0.586	0.503
R-squared	0.692	0.686	0.697	0.690	0.683	0.684	0.683	0.685	0.688
Observations	102	102	100	102	102	102	102	96	96
Number of clusters	50	50	49	50	50	50	50	47	47
Panel B: Panel	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
Baseline predicted mortality	-0.233	-0.215	-0.334	-0.230	-0.245	-0.248	-0.246	-0.206	-0.175
p-value	(0.012)	(0.005)	(0.008)	(0.008)	(0.003)	(0.003)	(0.003)	(0.017)	(0.032)
p-value, wild bootsrap	(0.002)	(0.008)	(0.010)	(0.002)	(0.002)	(0.006)	(0.002)	(0.008)	(0.016)
p-value for post year dummies x …	0.962	0.634	0.0117	0.235	0.638	0.149	0.584	0.417	0.189
R-squared	0.451	0.466	0.491	0.478	0.451	0.451	0.452	0.472	0.485
Observations	313	313	268	265	313	313	313	284	284
Number of clusters	63	63	53	52	63	63	63	57	57

Note.—OLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A are long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B are unbalanced panels with one observation per decade. See the text and Appendix Table 1 for definitions and details.

Appendix Table 6. Reduced Forms Robustness III: Population and Predicted Mortality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable is years in conflict/total years according to COW							
	Interaction of post-year dummies with…							
	Share of GDP in Natural Resource Sector in 1970	Share of Mineral Production in GNP in 1971	Oil Production in 1960	Diamond Production in 1960	Religious Polarization	Religious Fragmentation	Average Ethnolinguistic Fragmentation	Share of Catholic, Muslim, Protestant in 1980
Panel A: long differences	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980	Just 1940 and 1980
Baseline Predicted Mortality	-0.755	-0.536	-0.513	-0.517	-0.518	-0.531	-0.472	-0.615
p-value	(0.028)	(0.026)	(0.029)	(0.029)	(0.094)	(0.049)	(0.090)	(0.032)
p-value, wild bootsrap	(0.012)	(0.002)	(0.006)	(0.004)	(0.056)	(0.016)	(0.046)	(0.014)
p-value for post year dummy x …	0.142	0.457	0.502	0.131	0.990	0.870	0.784	0.308
R-squared	0.716	0.682	0.683	0.684	0.681	0.681	0.684	0.714
Observations	68	92	102	102	96	96	102	100
Number of clusters	33	45	50	50	47	47	50	49
Panel B: panel regressions	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980	Panel 1940-1980
Baseline Predicted Mortality	-0.384	-0.237	-0.246	-0.248	-0.199	-0.219	-0.197	-0.238
p-value	(0.009)	(0.007)	(0.003)	(0.003)	(0.063)	(0.025)	(0.005)	(0.015)
p-value, wild bootsrap	(0.004)	(0.002)	(0.002)	(0.002)	(0.038)	(0.016)	(0.004)	(0.006)
p-value for post year dummies x …	0.392	0.514	0.759	0.150	0.612	0.675	0.393	0.515
R-Squared	0.471	0.472	0.452	0.451	0.472	0.468	0.460	0.480
Observations	195	274	313	313	284	284	298	308
Number of clusters	39	55	63	63	57	57	60	62

Note.—OLS regressions with a full set of year and country fixed effects. Regular p-values for t-statistics with robust standard errors (clustered by country) are reported in parentheses, as well as cluster robust p-values following the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008). Panel A are long-difference specifications with two observations per country, one for the intial date and one for the final date. Panel B are unbalanced panels with one observation per decade. See the text and Appendix Table 1 for definitions and details.

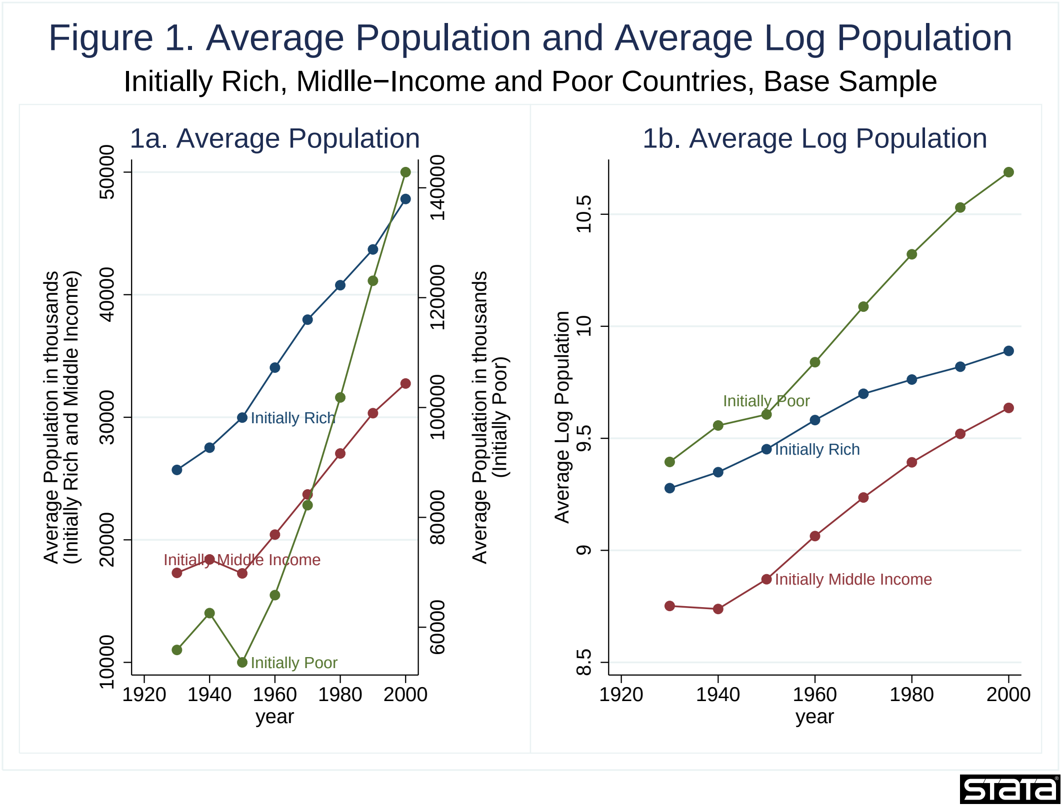
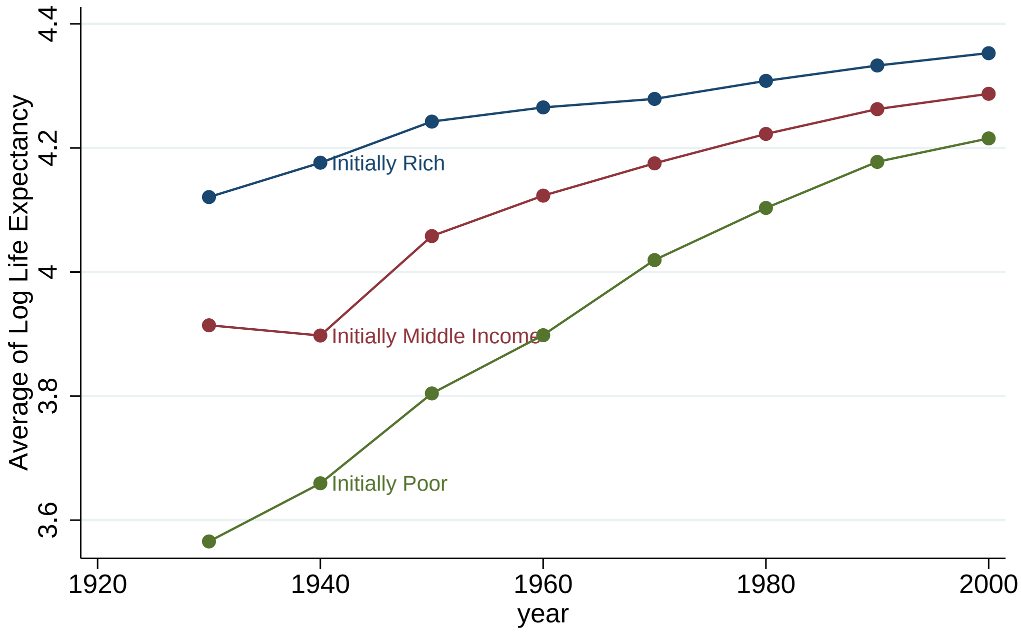


Figure 1. Average Population and Average Log Population
Initially Rich, Midlle−Income and Poor Countries, Base Sample
1a. Average Population
Average Population in thousands
(Initially Rich and Middle Income)
Average Population in thousands
(Initially Poor)
10000
20000
30000
40000
50000
60000
80000
100000
120000
140000
Initially Rich
Initially Middle Income
Initially Poor
1920
1940
1960
1980
2000
year
1b. Average Log Population
Average Log Population
8.5
9
9.5
10
10.5
Initially Poor
Initially Rich
Initially Middle Income
1920
1940
1960
1980
2000
year

Figure 2. Average Log Life Expectancy

Initially Rich, Midlle−Income and Poor Countries, Base Sample



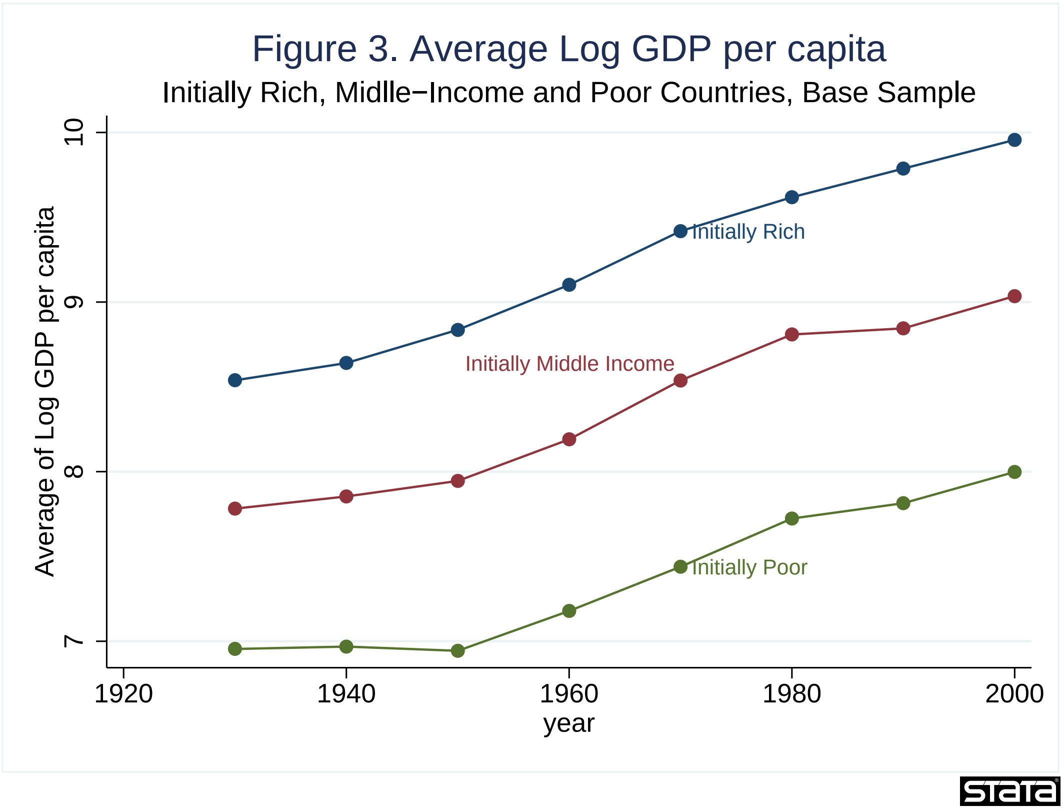


Figure 3. Average Log GDP per capita
Initially Rich, Midlle−Income and Poor Countries, Base Sample
Average of Log GDP per capita
10
9
8
7
1920
1940
1960
1980
2000
year
Initially Rich
Initially Middle Income
Initially Poor

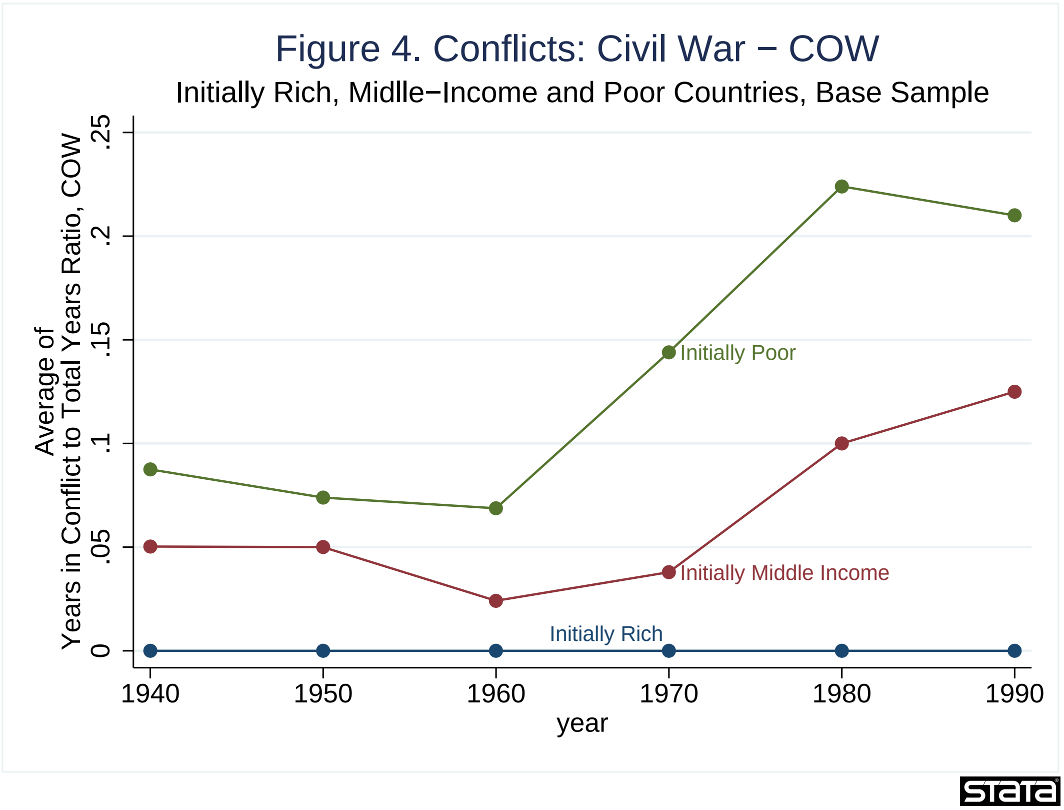


Figure 4. Conflicts: Civil War − COW
Initially Rich, Midlle−Income and Poor Countries, Base Sample
Average of
Years in Conflict to Total Years Ratio, COW
0
.05
.1
.15
.2
.25
Initially Poor
Initially Middle Income
Initially Rich
1940
1950
1960
1970
1980
1990
year

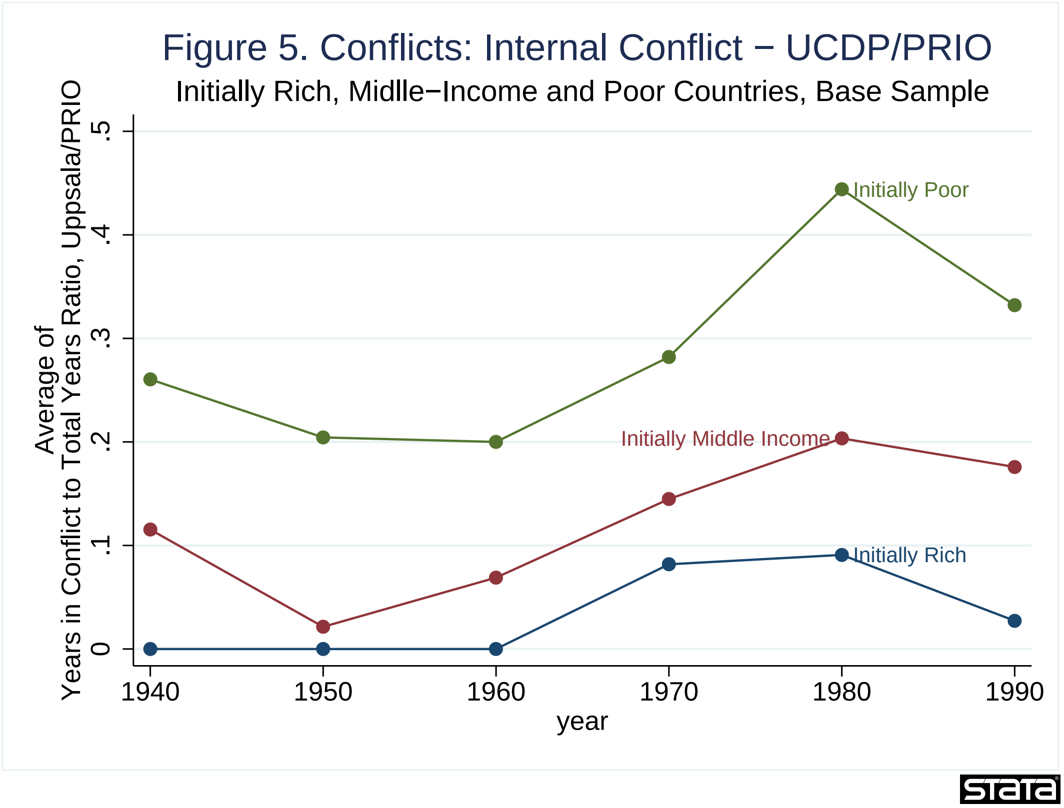


Figure 5. Conflicts: Internal Conflict − UCDP/PRIO
Initially Rich, Midlle−Income and Poor Countries, Base Sample
Average of
Years in Conflict to Total Years Ratio, Uppsala/PRIO
0
.1
.2
.3
.4
.5
1940
1950
1960
1970
1980
1990
year
Initially Poor
Initially Middle Income
Initially Rich

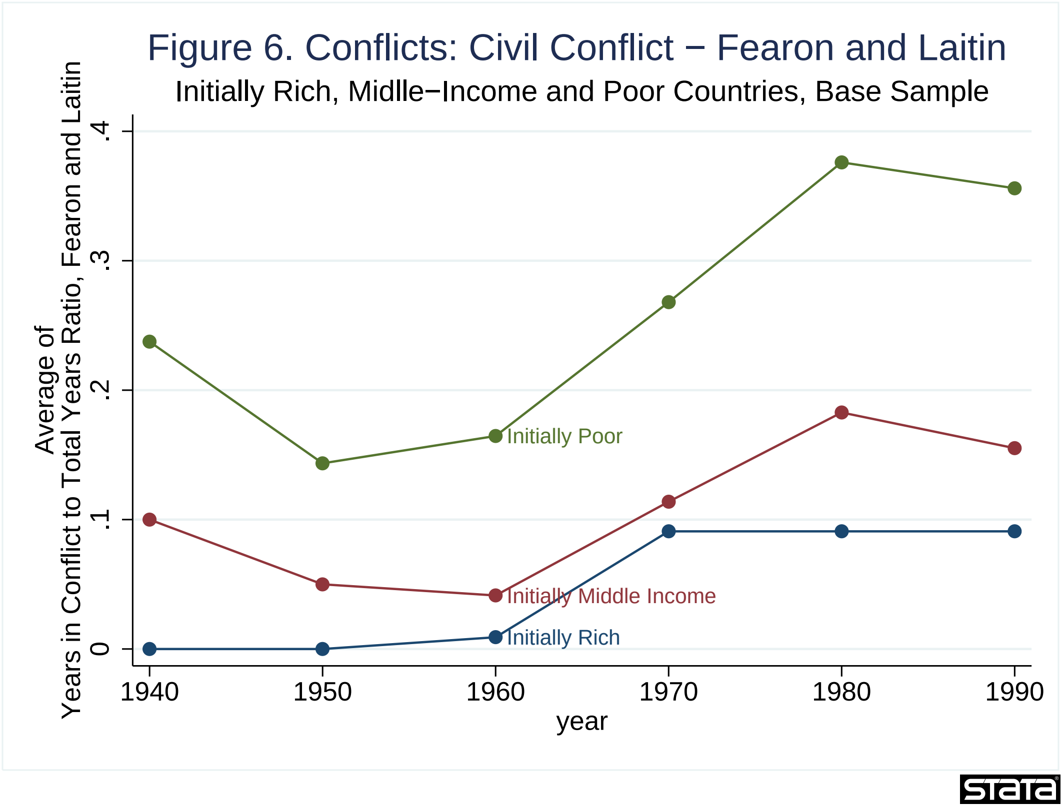


Figure 6. Conflicts: Civil Conflict − Fearon and Laitin
Initially Rich, Midlle−Income and Poor Countries, Base Sample
Average of
Years in Conflict to Total Years Ratio, Fearon and Laitin
0
.1
.2
.3
.4
1940
1950
1960
1970
1980
1990
year
Initially Poor
Initially Middle Income
Initially Rich

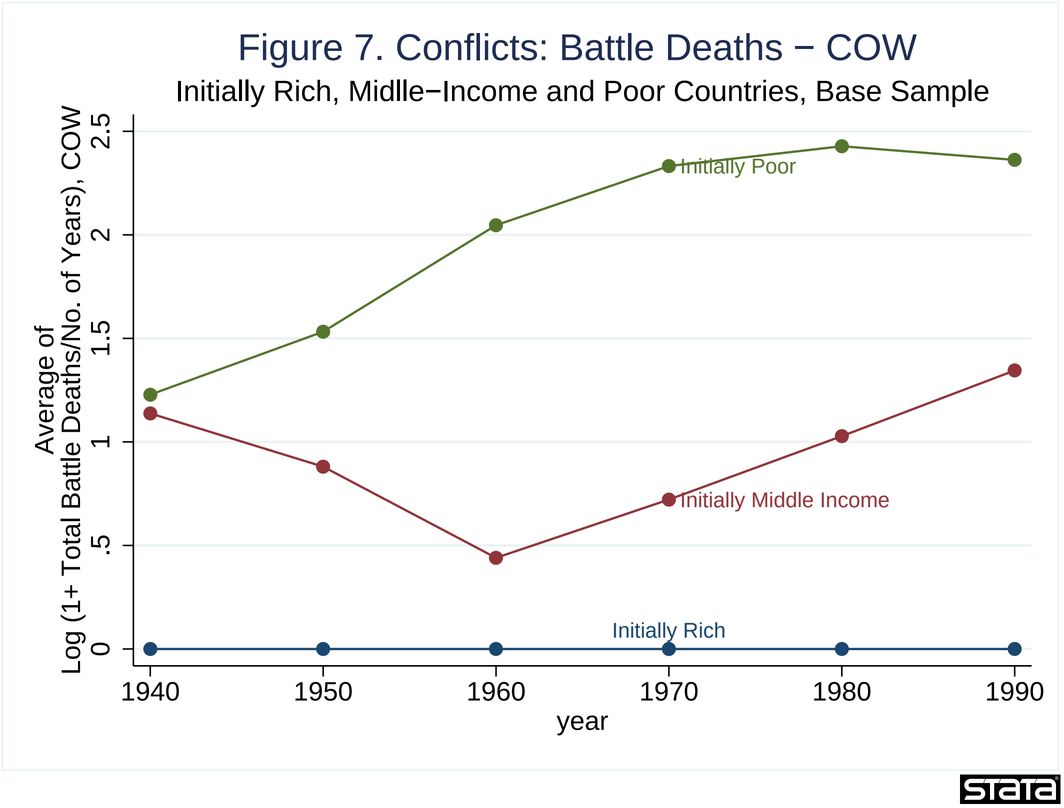


Figure 7. Conflicts: Battle Deaths − COW
Initially Rich, Midlle−Income and Poor Countries, Base Sample
Average of
Log (1+ Total Battle Deaths/No. of Years), COW
0
.5
1
1.5
2
2.5
Initially Poor
Initially Middle Income
Initially Rich
1940
1950
1960
1970
1980
1990
year

Figure 8
Change in Log Life Expectancy and Change in Predicted Mortality
Base Sample

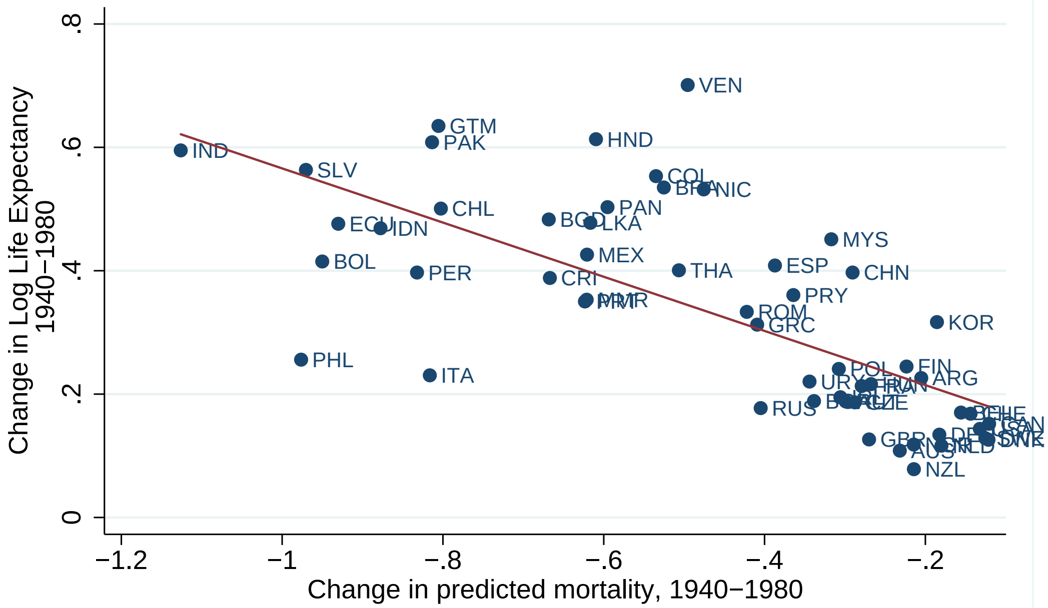


Figure 9
Change in Log Life Expectancy and Change in Predicted Mortality
Low and Middle Income Countries

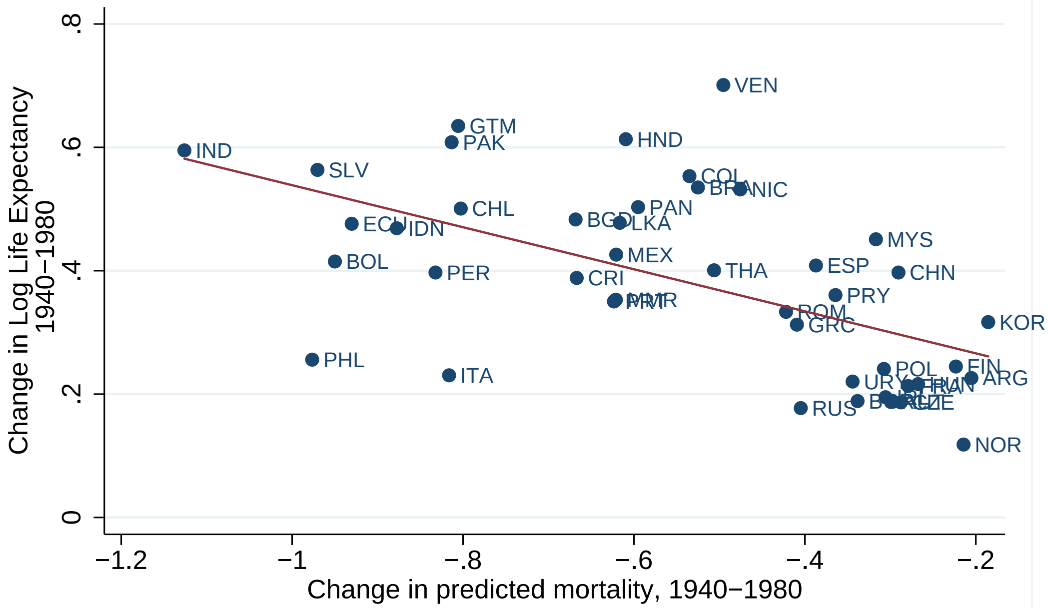


Figure 10
Change in Log Life Expectancy 1900−1940, and
Change in Predicted Mortality 1940−1980

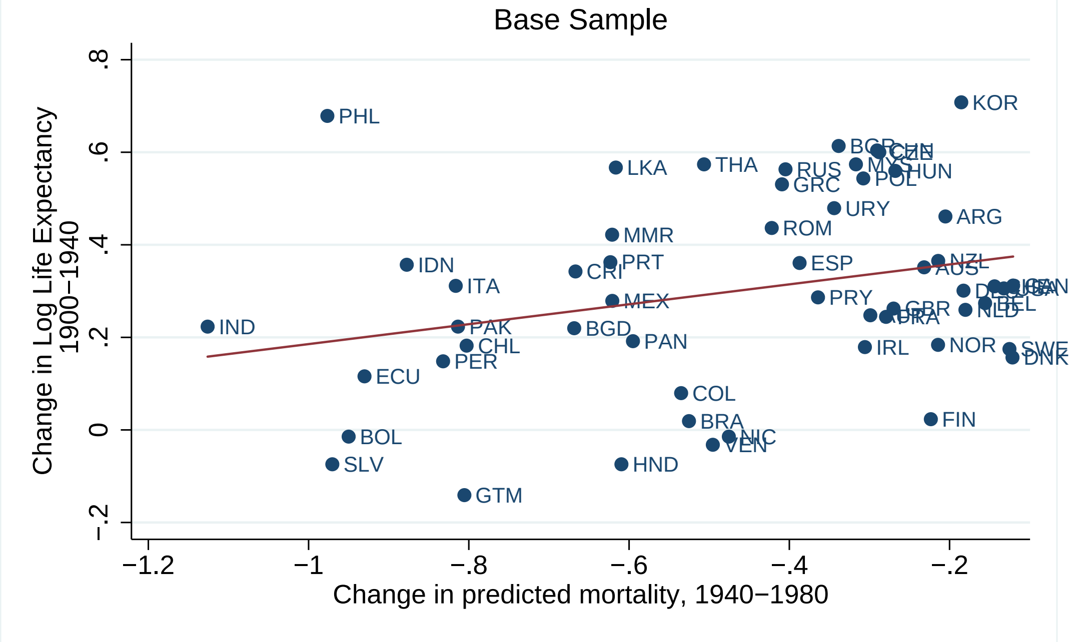


Figure 11
Change in Log Life Expectancy 1930−1940, and
Change in Predicted Mortality 1940−1980

Base Sample

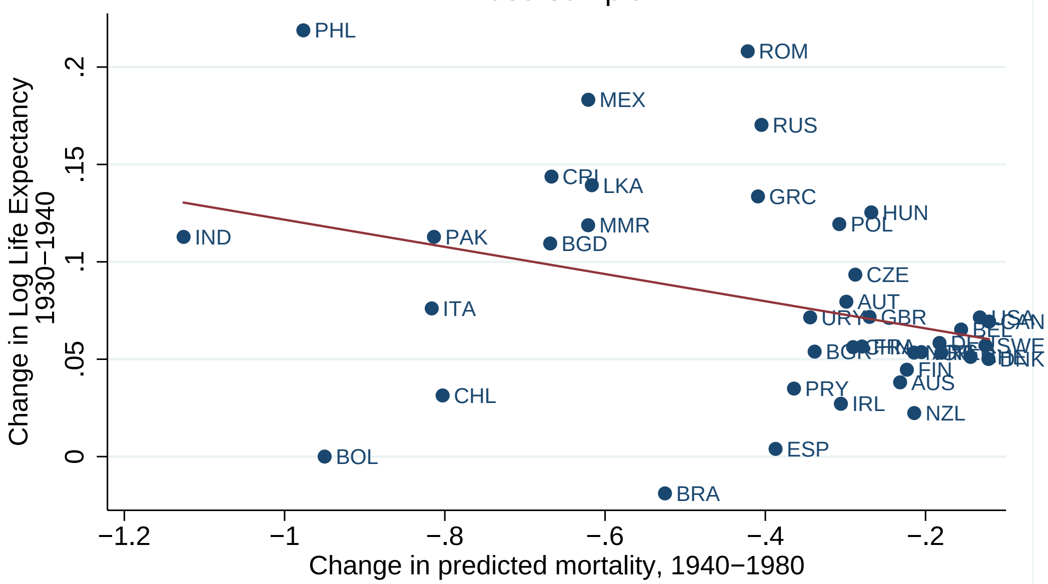
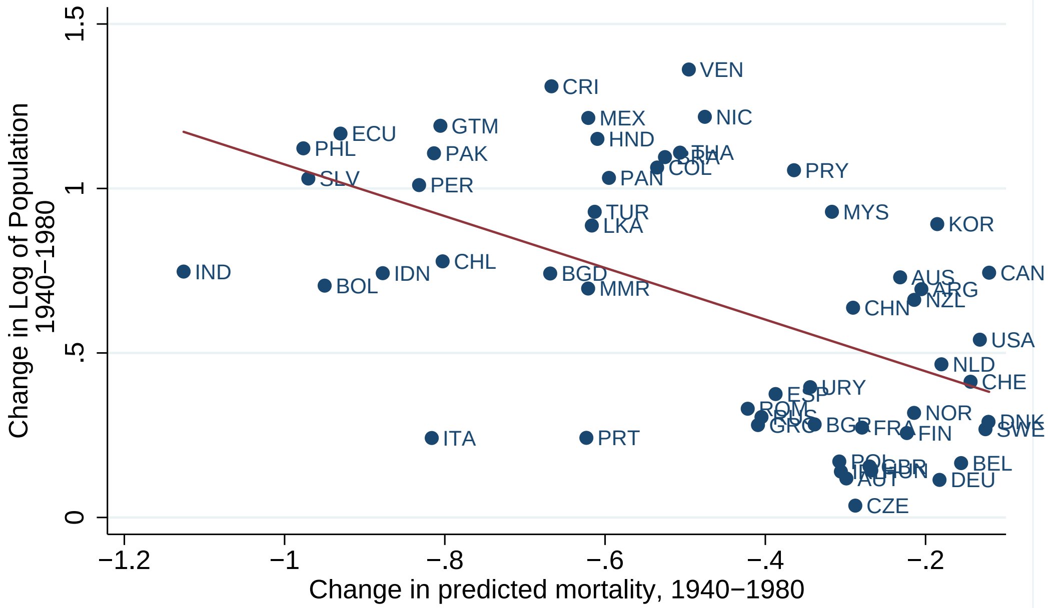


Figure 12
Change in Log Population and Change in Predicted Mortality
Base Sample



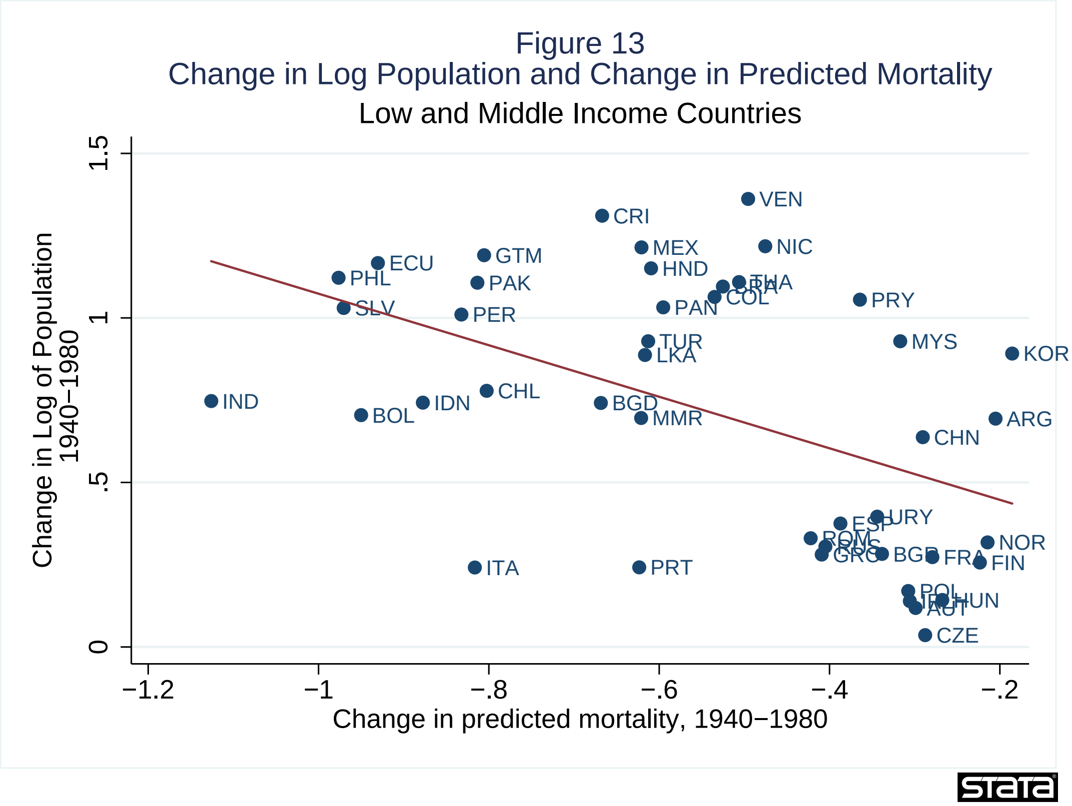


Figure 13
Change in Log Population and Change in Predicted Mortality
Low and Middle Income Countries
Change in Log of Population
1940−1980
Change in predicted mortality, 1940−1980
0
.5
1
1.5
−1.2
−1
−.8
−.6
−.4
−.2
VEN
CRI
MEX
NIC
ECU
GTM
HND
PHL
PAK
THA
COL
PRY
SLV
PER
PAN
TUR
MYS
KOR
LKA
IND
CHL
IDN
BGD
BOL
MMR
ARG
CHN
URY
ESP
ROM
NOR
GRC
BGR
FRA
FIN
ITA
PRT
POL
HUN
AUT
CZE

Figure 14
Change in Years in Conflict/Total Years and
Change in Predicted Mortality

Base Sample

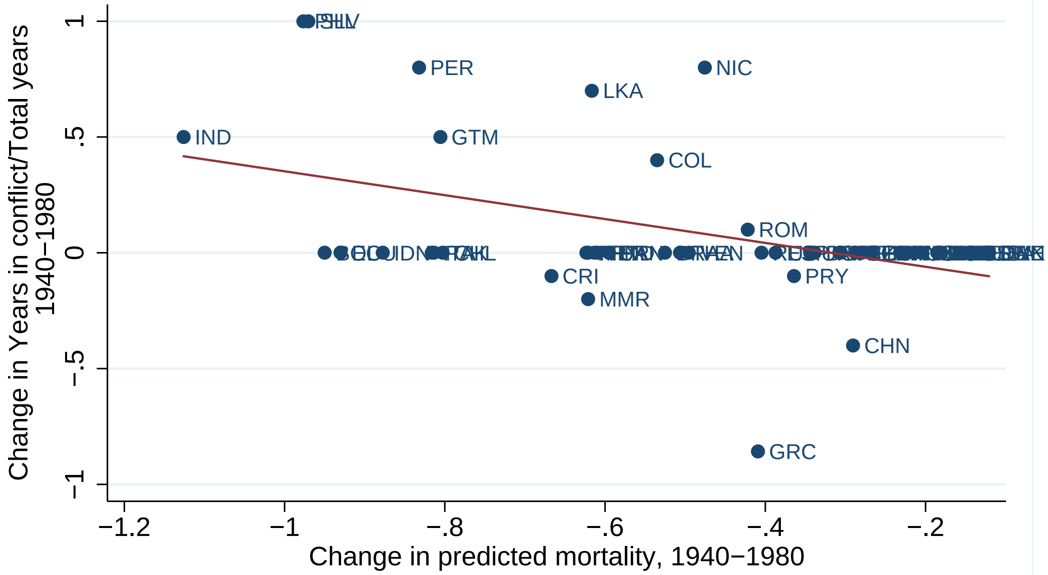


Figure 15
Change in Years in Conflict/Total Years and
Change in Predicted Mortality

Low and Middle Income Countries

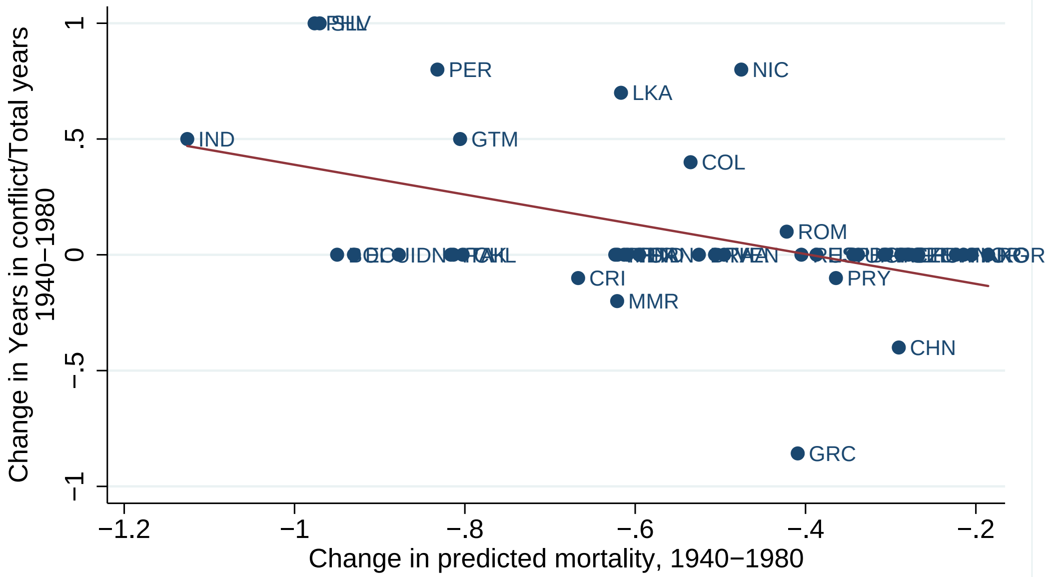


Figure 16
Change in Years in Conflict/Total Years 1900−1940, and
Change in Predicted Mortality 1940−1980

Base Sample

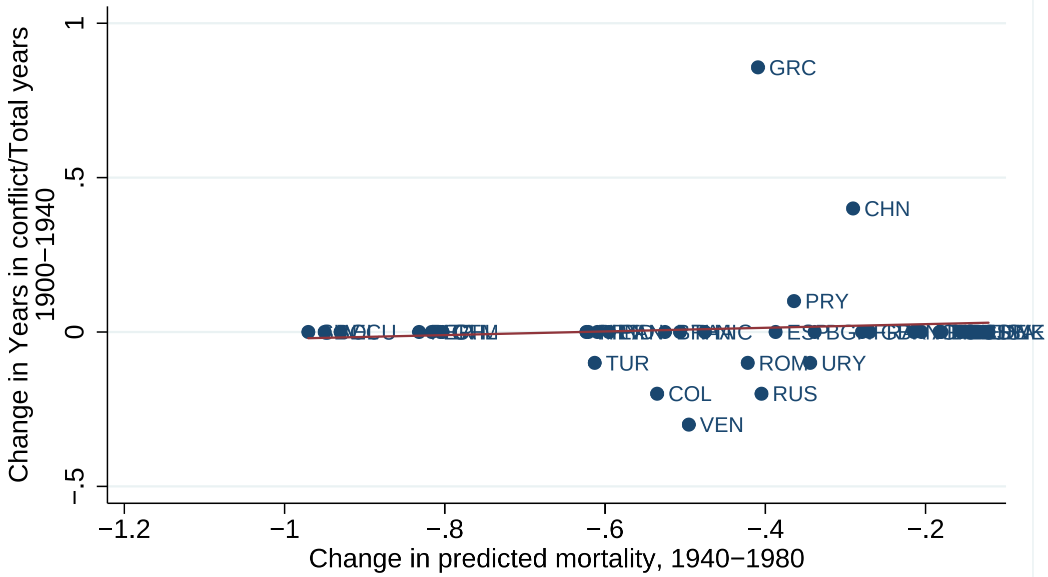


Figure 17
Change in Years in Conflict/Total Years 1900−1940, and
Change in Predicted Mortality 1940−1980

Low and Middle Income Countries