Economic Shocks and Civil Conflict: An Instrumental Variables Approach Data Set

Edward Miguel
National Bureau of Economic Research (NBER) and
Department of Economics
University of California, Berkeley
(emiguel@econ.berkeley.edu)

Shanker Satyanath
Department of Politics
New York University
(shanker.satyanath@nyu.edu)

Ernest Sergenti
Department of Politics
New York University
(ejs210@nyu.edu)

1. OVERVIEW

There is a growing body of research that highlights the association between economic conditions and civil conflict (see Sambanis 2001 for a review). The existing literature, however, does not adequately address the endogeneity of economic variables to civil war, and thus does not convincingly establish a causal relationship. In addition to endogeneity, omitted variables – for example, government institutional quality – may drive both economic outcomes and conflict, producing misleading cross-country estimates.

In "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," we use exogenous variation in rainfall as an instrumental variable for income growth in order to estimate the impact of economic growth on civil conflict. Weather shocks are plausible instruments for GDP growth in economies that largely rely on rain-fed agriculture, i.e., neither have extensive irrigation systems nor are heavily industrialized. The instrumental variable method makes it credible to assert that the association between economic conditions and civil war is a causal relationship, rather than simply a correlation.

Sub-Saharan Africa is the ideal region for this identification strategy: the World Development Indicator (WDI) database indicates that only one percent of crop land is irrigated in the median African country, and the agricultural sector remains large.

The data used in "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," are of four kinds: rainfall; conflict; economic, demographic, and development controls; and political institutional controls. The most original of

these four is the rainfall data. Therefore, we devote an entire section of this manual, section 2, to a description of our four rainfall measures and an explanation of the methodology used to construct each. In section 3, we describe all of our data sources by category of data and we provide links to where the original information may be found on-line. A detailed description of each variable in the set is provided in section 4. In addition, this manual contains two appendices: appendix A lists all latitude and longitude points used to generate two of the four rainfall measures, the GPCP and NCEP measures, and appendix B presents the Stata codebook for all variables in the dataset.

2. RAINFALL DATA & CONSTRUCTION OF THE RAINFALL MEASURES

We employ four rainfall data sets:

A. Global Precipitation Climatology Project (GPCP)

http://cics.umd.edu/GPCP

B. National Centers for Environment Prediction (NCEP)

http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.Merged_Analysis/.monthly/

C. U.N. Food and Agricultural Organization Climatic (FAOCLIM2) Data CD-ROM: World-Wide Agroclimatic Database. FAO-Agrometeorology Group, Rome 2000.

D. Normalized Difference Vegetation Index (NDVI)

http://edcw2ks21.cr.usgs.gov/adds/

A. Global Precipitation Climatology Project (GPCP) Data Set

The Global Precipitation Climatology Project (GPCP) database of rainfall estimates stretches back to 1979. The GPCP data rely on a combination of actual weather station rainfall gauge measures and satellite information on the density of cold cloud cover, which is closely related to actual precipitation. The GPCP uses the Huffman et al. (1995, 1997) method of data selection and merging.

Estimates are made at 2.5 latitude and longitude degree intervals. The units of measurement are in millimeters of rainfall per day and are the average

per month. We multiply each monthly average by the number of days in a given month, which gives us an estimate of total monthly rainfall.[1] We then add up all of the total monthly estimates in a given year to generate an estimate of total yearly rainfall for each 2.5 latitude / longitude degree node. For example, the yearly rainfall estimate for any 2.5 latitude / longitude degree node in 1999 was calculated as follows:

y1999 = a9901*31 + a9902*28 + a9903*31 + a9904*30 + a9905*31 + a9906*30 + a9907*31 + a9908*31 + a9909*30 + a9910*31 + a9911*30 + a9912*31

where aYYMM is the average daily rainfall in millimeters for month MM and year YY taken from the GPCP data set

Next, each yearly rainfall estimate per 2.5 latitude / longitude degree node is averaged over all nodes in a given country to produce an estimate of total yearly rainfall per country. For example, our estimate of total yearly rainfall for Kenya is the average of the yearly rainfall estimates for the eight 2.5 latitude / longitude degree nodes in Kenya. See Appendix A for a listing of all nodes used in the calculation of the each country's rainfall estimates.

(Note: No degree grid node fell within the national boundaries for five small African countries – Burundi, Djibouti, Gambia, Guinea-Bissau, and Rwanda. In these cases, we assigned the rainfall measures from the nearest node(s) to their borders. See Appendix A.)

[1] Note that the following years contained a leap year: 1980, 1984, 1988, 1992, 1996, and 2000. For these years, we multiplied the average daily rainfall for February by 29 days instead of 28 days.

B. National Centers for Environment Prediction (NCEP) Data Set

This data set is essentially similar to the GPCP data set presented above. It differs in that it uses the Xie and Arkin (1997) method of data selection and merging. The construction of the total yearly estimates per country is identical to the one used with the GPCP.

C. U.N. Food and Agricultural Organization Climatic (FAOCLIM2) Data

The FAOCLIM2 data set relies solely on gauge measures. Data are available starting in the early 1800's for some countries. Unfortunately, rain gauge coverage becomes increasingly limited after 1990, and especially after 1996, leading to missing observations.

The units of measurement are in millimeters of rainfall per month per gauge station. We first calculate the average rainfall per month for the entire country by taking the average of the rainfall per month measurements across gauge stations. We then add up all of the country monthly averages in a given year to generate our measure of total yearly rainfall per country.
(Note: It is often the case that data are not available for many gauge stations. That is, the total number of gauge stations used to calculate the average rainfall per month is not constant. Therefore, we include an additional variable in our data set, sm_obs, the total number of station-month FAOCLIM2 observations per

year per country, to provide an estimate of the degree of precision of the total yearly rainfall per country measure.)

D. Normalized Difference Vegetation Index (NDVI) Data Set

The Normalized Difference Vegetation Index (NDVI) database of rainfall estimates also stretches back to 1979. NDVI provides a measure of the living green plant biomass on the surface of the Earth. It differs, therefore, from the other measures that employ satellite imaging, GPCP and NCEP, in that it estimates vegetation on the Earth and not the density of cold cloud cover. Nevertheless, NDVI is closely related to rainfall, with a correlation of 0.9. We do not actively employ this measure in our paper, however, as vegetation levels may be a function of crop choices made in response to civil conflict, and thus could be endogenous to the conflict.

The methodology used to construct the total yearly estimates per country is similar to the one used with GPCP and NCEP. The major difference is, with NDVI, we use latitude and longitude intervals of 1 degree and estimates are made per dekad (roughly 10 days), whereas, with GPCP and NCEP, we use latitude and longitude intervals of 2.5 degrees and estimates are made per month.

Like before, first, we add up all of the total dekad estimates in a given year to generate an estimate of total yearly rainfall for each 1 degree latitude / longitude node. And, next, each yearly rainfall estimate per 1 degree latitude /

longitude node is averaged over all nodes in a given country to produce an estimate of total yearly rainfall per country.

Finally, note, in our paper we focus on the GPCP dataset over the other four. First, as mentioned above, NDVI may suffer from endogenity. Second, of the three remaining sources, GPCP is the only one that at the same time: includes both gauge and satellite data; corrects for systematic errors in gauge measures; and rejects gauge measures thought to be unreliable (Rudolf 2000).

3. THE MSS Civil War Data Set - Overview

The MSS Civil War Data Set combines data from several sources. The data sets used are listed below, by category of data: rainfall; civil conflict; economic, demographic and development controls; and political institutional controls.

(Note: some sets fall into several categories. When this occurs, the set is listed only once under the first category on our list. For example, from the Fearon and Laitin set, we obtained civil war, economic, and political institutional data. This set is listed under the civil war category.)

1. Rainfall Data Sets

A. Global Precipitation Climatology Project (GPCP)
   http://cics.umd.edu/GPCP

B. National Centers for Environment Prediction (NCEP)
   http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.Merged_Analysis/.monthly/

C. U.N. Food and Agricultural Organization Climatic (FAOCLIM2) Data
   CD-ROM: World-Wide Agroclimatic Database.
   FAO-Agrometeorology Group, Rome 2000.

D. Normalized Difference Vegetation Index (NDVI)
   http://edcw2ks21.cr.usgs.gov/adds/

2. Civil Conflict Data Sets

A. Armed Conflict Data – International Peace Research Institute of Oslo, Norway and the University of Uppsala, Sweden (PRIO/Uppsala)

http://www.prio.no/cwp/ArmedConflict

B. Fearon and Laitin (2003)

American Political Science Review, 97(1), 75-90.

http://www.stanford.edu/group/ethnic/publicdata/publicdata.html

C. Doyle and Sambanis

(from Sambanis - Journal of Conflict Resolution vol. 45, no. 3)

http://www.yale.edu/unsy/civilwars/data.htm

3. Economic, Demographic, and Development Controls

A. Global Development Network Growth Database (GDNGD)

http://www.nyu.edu/fas/institute/dri/index.html

B. World Development Indicators (WDI)

CD-ROM: World Development Indicators 2002 Database.

Washington, D.C. 2002.

C. FAO Stat

http://apps.fao.org/default.jsp

D. Fractionalization Data

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio

Kurlat, and Romain Wacziarg - Journal of Economic Growth, vol. 8, no. 2, 155-194.

http://www.stanford.edu/~wacziarg/papersum.html

4. Political Institutional Controls

A. Polity IV

http://www.cidcm.umd.edu/inscr/polity/

B. Database of Political Institutions (DPI)

http://www.worldbank.org/research/bios/pkeefer.htm

C. Freedom House

http://www.freedomhouse.org/ratings/index.htm

D. *The Logic of Political Survival* Data Set

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2003. Cambridge: MIT Press.

http://www.nyu.edu/gsas/dept/politics/data/bdm2s2/Logic.htm

E. Barro (1991)

(from Levine and Renelt – American Economic Review LXXXII (1992), 942-963.)

http://www.worldbank.org/research/growth/ddlevren.htm

4. THE MSS Civil War Data Set – Variable Descriptions

0. Identification Variables

CCODE
Correlates of War (COW) Country Code

YEAR_ACTUAL
From 1981 or the first year of independence of the country

COUNTRY_NAME
Country Name

COUNTRY_CODE
Secondary Country Code. Often but not always the same as the World Bank Country Code

1. Rainfall Variables

A. Global Precipitation Climatology Project (GPCP) and Derived

GPCP
Global Precipitation Climatology Project estimate of average precipitation in millimeters per year. The exact source was NASA GPCP V2. It uses the Huffman et al. (1995, 1997) method of data selection and merging. See section 2 for an explanation of the methodology used to construct this measure.
Source: Global Precipitation Climatology Project (GPCP)

GPCP_L
GPCP lagged one year

GPCP_L2
GPCP lagged two years

GPCP_G
GPCP growth: (GPCP - GPCP_l) / (GPCP_l)

GPCP_G_L
GPCP growth lagged one year: (GPCP_l - GPCP_l2) / (GPCP_l2)

GPCP_G_FL
GPCP growth lagged forward one year:

GPCP_g_fl = GPCP_g[_n+1] if ccode==ccode[_n+1]

GPCP_D
GPCP first difference: (GPCP – GPCP_l)

GPCP_D_L
GPCP first difference lagged one year: (GPCP_l – GPCP_l2) / (GPCP_l2)

GPCP_DF_MEAN
GPCP difference from the mean (of the 1979 to 2001 observations)

GPCP_DF_MEAN_1
GPCP_df_mean lagged one year

GPCP_DF_MEAN_2
GPCP_df_mean lagged two years

B. National Centers for Environment Prediction (NCEP) and Derived

NCEP
National Centers for Environment Prediction (NCEP) estimate of average precipitation in millimeters per year. The exact source was NOAA NCEP CPC Merged Analysis. It uses the Xie and Arkin (1997) method of data selection and merging. See section 2 for an explanation of the methodology used to construct this measure.
Source: National Centers for Environment Prediction (NCEP)

NCEP_L
NCEP lagged one year

NCEP_L2
NCEP lagged two years

NCEP_G
NCEP growth: (NCEP - NCEP_l) / (NCEP_l)

NCEP_G_L
NCEP growth lagged one year: (NCEP_l - NCEP_l2) / (NCEP_l2)

NCEP_G_FL
NCEP growth lagged forward one year:
NCEP_g_fl = NCEP_g[_n+1] if ccode==ccode[_n+1]

NCEP_D
NCEP first difference: (NCEP – NCEP_l)

NCEP_D_L
NCEP first difference lagged one year: (NCEP_l – NCEP_l2) / (NCEP_l2)

NCEP_DF_MEAN
NCEP difference from the mean (of the 1979 to 2001 observations)

NCEP_DF_MEAN_1
NCEP_df_mean lagged one year

NCEP_DF_MEAN_2
NCEP_df_mean lagged two years

C. U.N. FAO Climatic (FAOCLIM2) Database and Derived

SM_OBS
Number of Station-Month observations used in the calculation of the FAO average precipitation in millimeters per year index.
Source: FAOCLIM2

FAO
FAO Climatic (FAOCLIM2) Database estimate of average precipitation in millimeters per year. See section 2 for an explanation of the methodology used to construct this measure.
Source: FAOCLIM2

FAO_L
FAO lagged one year

FAO_L2
FAO lagged two years

FAO_G
FAO growth: (FAO - FAO_l) / (FAO_l)

FAO_G_L
FAO growth lagged one year: (FAO_l - FAO_l2) / (FAO_l2)

FAO_G_FL
FAO growth lagged forward one year:
FAO_g_fl = FAO_g[_n+1] if ccode==ccode[_n+1]

FAO_D
FAO first difference: (FAO – FAO_l)

FAO_D_L
FAO first difference lagged one year: (FAO_l – FAO_l2) / (FAO_l2)

FAO_DF_MEAN
FAO difference from the mean (of the 1960 to the latest available observations)

FAO_DF_MEAN_1
FAO_df_mean lagged one year

FAO_DF_MEAN_2
FAO_df_mean lagged two years

D. Normalized Difference Vegetation Index (NDVI) and Derived

NDVI
Normalized Difference Vegetation Index (NDVI) estimate of the density of plant life, closely related to rainfall in Africa. See section 2 for an explanation of the methodology used to construct this measure.
Source: Normalized Difference Vegetation Index (NDVI)

NDVI_L
NDVI lagged one year

NDVI_L2
NDVI lagged two years

NDVI_G
NDVI growth: (NDVI - NDVI_l) / (NDVI_l)

NDVI_G_L
NDVI growth lagged one year: (NDVI_l - NDVI_l2) / (NDVI_l2)

NDVI_G_FL
NDVI growth lagged forward one year:
NDVI_g_fl = NDVI_g[_n+1] if ccode==ccode[_n+1]

NDVI_D
NDVI first difference: (NDVI – NDVI_l)

NDVI_D_L
NDVI first difference lagged one year: (NDVI_l – NDVI_l2) / (NDVI_l2)

NDVI_DF_MEAN
NDVI difference from the mean (of the 1982 to 2001 observations)

NDVI_DF_MEAN_1
NDVI_df_mean lagged one year

NDVI_DF_MEAN_2
NDVI_df_mean lagged two years

2. Civil War Variables

A. PRIO/Uppsala Armed Conflict Data (Monadic) and Derived

TYPE3
Type3 is PRIO/Uppsala's indicator of Internal Conflict. It can take on four distinct values:
0: No Internal Conflict
1: Internal Minor Armed Conflict
2: Internal Intermediate Armed Conflict
3: Internal War

PRIO/Uppsala define Minor Conflict, Intermediate Conflict, and War as follows:
• Minor Armed Conflict: At least 25 battle-related deaths per year and fewer than 1,000 battle-related deaths during the course of the conflict.
• Intermediate Armed Conflict: At least 25 battle-related deaths per year and an accumulated total of at least 1,000 deaths, but fewer than 1,000 per year.
• War: At least 1,000 battle-related deaths per year.

Source: PRIO/Uppsala Armed Conflict Data

TYPE4
Type4 is PRIO/Uppsala's indicator of Internationalized Internal Conflict. It can take on four distinct values:
0: No Internationalized Internal Conflict
1: Internationalized Internal Minor Armed Conflict
2: Internationalized Internal Intermediate Armed Conflict
3: Internationalized Internal War

PRIO/Uppsala define Minor Conflict, Intermediate Conflict, and War as follows:
• Minor Armed Conflict: At least 25 battle-related deaths per year and fewer than 1,000 battle-related deaths during the course of the conflict.
• Intermediate Armed Conflict: At least 25 battle-related deaths per year and an accumulated total of at least 1,000 deaths, but fewer than 1,000 per year.
• War: At least 1,000 battle-related deaths per year.

Source: PRIO/Uppsala Armed Conflict Data

WAR_PRIO
Internal War or Internationalized Internal War.
Internal Conflict or Internationalized Conflict with at least 1,000 battle-related deaths per year. Dichotomous variable. Coded “1” if TYPE3 equals 3 or TYPE4 equals 3, “0” otherwise.

MINOR_PRIO
Minor or Intermediate Internal Conflict or Minor or Intermediate Internationalized Conflict. Dichotomous variable. Coded “1” if TYPE3 equals 1 or 2 or TYPE4 equals 1 or 2, “0” otherwise.

ANY_PRIO
Any Internal War or Any Internationalized Internal War.
Dichotomous variable. Coded “1” if TYPE3 equals 1, 2, or 3 or TYPE4 equals 1, 2, or 3, “0” otherwise.

WAR_PRIO_ON
WAR_PRIO Onset. Dichotomous variable. Coded “1” if Internal War or Internationalized Internal War onset during country year, “0” otherwise.

MINOR_PRIO_ON
MINOR_PRIO Onset. Dichotomous variable. Coded “1” if Minor or Intermediate Internal Conflict or Minor or Intermediate Internationalized Conflict onset during country year, “0” otherwise.

ANY_PRIO_ON
ANY_PRIO Onset. Dichotomous variable. Coded “1” if Any Internal War or Any Internationalized Internal War onset during country year, “0” otherwise.

WAR_PRIO_OFF
WAR_PRIO Offset. Dichotomous variable. Coded “1” if Internal War or Internationalized Internal War ends during country year, “0” otherwise.

MINOR_PRIO_OFF
MINOR_PRIO Offset. Dichotomous variable. Coded “1” if Minor or Intermediate Internal Conflict or Minor or Intermediate Internationalized Conflict ends during country year, “0” otherwise.

ANY_PRIO_OFF
ANY_PRIO Offset. Dichotomous variable. Coded “1” if Any Internal War or Any Internationalized Internal War ends during country year, “0” otherwise.

B. Fearon and Laitin (2003) and Derived

WARS
Number of wars in progress during a given country year
Source: Fearon and Laitin (2003)

WAR
Dichotomous variable. Coded "1" if war ongoing during country year, "0" otherwise.
Source: Fearon and Laitin (2003)

WARL
WAR lagged one year, with 0 for start of country series.
Source: Fearon and Laitin (2003)

ONSET
Onset of a Civil War. Dichotomous variable. Coded "1" if civil war onset during country year, "0" otherwise.
Source: Fearon and Laitin (2003)

ENDED
End of Civil War. Coded "1" if civil war ends during country year, "0" if ongoing, Missing(.) otherwise.
Source: Fearon and Laitin (2003)

WARYRS
Number of War Years for each onset
Source: Fearon and Laitin (2003)

POP
Population, in thousands.
"For the country years for which it is available, we used the Penn World Tables 5.6 numbers. Otherwise, we used the World Bank estimate (WDI 2001), and then the figure from the Correlates of War National Capabilities Data when neither the World Bank nor PWT provided an estimate. This means that population in years after 1992 are mainly World Bank estimates, while before 1950 everything is from COW. The correlation between these three different sources is nearly perfect, however, so it matters not all which source is used as the 'base.'"[2]
Source: Fearon and Laitin (2003)

LPOP
Log of pop
Source: Fearon and Laitin (2003)

---

[2] Fearon, James and David Laitin. "Additional Tables for 'Ethnicity, Insurgency and Civil War.'" Stanford University, February 6, 2003, p 3.

POLITY2
Revised polity score.
Taken from the Polity IV dataset. Polity is the difference between Polity IV's measure of democracy minus its measure of autocracy. Values range from –10 to 10. The revised polity score fills in missing values based on the following coding: when polity = -66, set polity2 = NULL, when polity = -77, set polity2 = 0, when polity = -88, extrapolate based previous and subsequent values.
Source: Fearon and Laitin (2003)

GDPEN
Per Capita GDP.
"We started with the Penn World Tables 5.6 for real per capita income (chain index), measured in 1985 U.S. dollars. This series starts in 1950 and ends in 1992, and provides estimates for 4,243 of our 6,610 country years (64%). We then used the estimates of growth rate of per capita income provided in the 2001 World Development Indicators (WDI, published by the World Bank) to extend these estimates forward to 1999 and backwards to the first year of independence or 1960 (the first year in the WDI data) where possible. This added another 1,116 observations (17% of country years)…."[3]
Source: Fearon and Laitin (2003)

GDPENL
GDPEN lagged one year, with 0 for start of country series.
Source: Fearon and Laitin (2003)

LGDPENL1
Log of GDPENL
Source: Fearon and Laitin (2003)

LPOPL1
Log of population lagged one year, with 0 for start of country series.
Source: Fearon and Laitin (2003)

COLBRIT
Dichotomous variable. Coded "1" if country was a former British colony, "0" otherwise.
Source: Fearon and Laitin (2003)

COLFRA
Dichotomous variable. Coded "1" if country was a former French colony, "0" otherwise.
Source: Fearon and Laitin (2003)

[3] Ibid, p 1.

MTNEST
Percent Mountainous Terrain. Based on work by geographer A.J. Gerard for the World Bank's "Economics of Civil War, Crime, and Violence" project.
Source: Fearon and Laitin (2003)

LMTNEST
Log of Mtnest
Source: Fearon and Laitin (2003)

OIL
Oil Exporters.
"We used World Bank (WDI) data on fuel exports as a percentage of merchandise exports, which is available for five year periods from 1960 and annually from 1980 for most countries. Missing years prior to 1980 and after 1960 were linearly interpolated where possible. We next created a dummy variable marking country years that had greater than 33% fuel exports."[4]
Source: Fearon and Laitin (2003)

NCONTIG
Noncontiguous State. Dichotomous variable. Coded "1" if a country is a non-continuous state, "0" otherwise.
Source: Fearon and Laitin (2003)

ETHFRAC
Ethnic-linguistic fractionalization based on the Atlas Marodov Mira.
Source: Fearon and Laitin (2003)

EF
Ethnic fractionalization based on Fearon (2002).
Source: Fearon and Laitin (2003)

RELFRAC
Religious Fractionalization.
"R. Quinn Mecham started with the CIA Factbook estimates and then used a number of other sources to construct a list of religions by country, and percentage of adherents. Figures are generally for the 1990s, though with few exceptions this variable does not seem to change much over time."[5]
Source: Fearon and Laitin (2003)

NWSTATE
New State. Dichotomous variable. Coded "1" if state is in its first two years of existence, "0" otherwise.
Source: Fearon and Laitin (2003)

---

[4] Ibid, p 4.
[5] Ibid, p 4.

POLITY2L
Polity2 lagged one year, with 0 for start of country series.
Source: Fearon and Laitin (2003)

INSTAB
Instablitiy. Greater than 2 change in Polity2 measure in last 3 years.
Source: Fearon and Laitin (2003)

DEML
Lagged Democracy. Dichotomous variable. Coded "1" if polity2l > 5, "0" otherwise.
Source: Fearon and Laitin (2003)

SDWARS
Number of Civil Wars in progress using Doyle and Sambanis's coding.
Source: Fearon and Laitin (2003)

SDONSET
Civil War Onset using Doyle and Sambanis's coding. Dichotomous variable. Coded "1" if onset in current year, "0" otherwise.
Source: Fearon and Laitin (2003)

COLWARS
Number of Civil Wars in progress using Collier and Hoeffler's coding.
Source: Fearon and Laitin (2003)

COLONSET
Civil War Onset using Collier and Hoeffler's coding. Dichotomous variable. Coded "1" if onset in current year, "0" otherwise.
Source: Fearon and Laitin (2003)

COWWARS
Number of Civil Wars in progress using the Correlates of War (COW) coding.
Source: Fearon and Laitin (2003)

COWONSET
Civil War Onset using the Correlates of War (COW) coding. Coded "1" if onset in current year, "0" otherwise.
Source: Fearon and Laitin (2003)

COWWARL
War in last period using Correlates of War (COW) coding. Coded "1" if COW war ongoing in last period.
Source: Fearon and Laitin (2003)

SDWARL
War in last period using Doyle and Sambanis. Dichotomous variable. Coded "1" if Doyle and Sambanis war ongoing in last period, "0" otherwise.
Source: Fearon and Laitin (2003)

COLWARL
War in last period using Collier and Hoeffler's coding. Dichotomous variable. Coded "1" if Collier and Hoeffler war ongoing in last period, "0" otherwise.
Source: Fearon and Laitin (2003)

GDP_G
GDP Growth. (GDPEN - GDPENL) / (GDPENL)

GDP_G_L
GDP_G lagged one year

Y_0
GDP per capita at the beginning of the period of analysis, 1979 (1990 for Namibia). GDPEN for 1979 (1990 for Namibia)

POLITY2L_6
Democracy Indicator. Dichotomous variable. Coded "1" if POLITY2L >= 6, "0" otherwise.

WAR_COL
Civil War Incidence using Collier and Hoeffler's coding. Dichotomous variable. Coded "1" if COLWARS > 0, "0" otherwise.

C. Doyle and Sambanis and Derived

WARSTDS
Civil War Start.
1: First Observation; Missing(.): Other observations of war; 0: No War
Source: Doyle and Sambanis

WAR_ON
Civil War Onset using Doyle and Sambanis coding. Dichotomous variable. Coded "1" if WARSTDS = 1, "0" otherwise.

WAR_INC
Civil War Incidence using Doyle and Sambanis coding. Dichotomous variable. Coded "1" if WARSTDS = 1 or WARSTDS = Missing(.), "0" otherwise.

3. Economic, Demographic, and Development Controls

A. Global Development Network Growth Database (GDNGD) and Derived

TOT_100
Terms of trade (goods and services, 1995 = 100)
Source: GDNGD

TOT_100_L
TOT_100 lagged one year

TOT_100_G
TOT_100 growth: (TOT_100 - TOT_100_l) / (TOT_100_l)

B. World Development Indicators (WDI) and Derived

TOT_ADJ
Terms of trade adjusted (constant Local Currency Units) - NY.TTF.GNFS.KN
Source: WDI

TRADE_PGDP
Trade as a percentage of GDP - NE.TRD.GNFS.ZS
Source: WDI

TRADE_GOODS_PGDP
Trade in goods as a percentage of GDP - TG.VAL.TOTL.GD.ZS
Source: WDI

TRADE_GOODS_PGOODSGDP
Trade in goods as a percentage of goods GDP - TG.VAL.TOTL.GG.ZS
Source: WDI

MIL_EXP
Military Expenditure as a percentage of central government expenditure - MS.MIL.XPND.ZS
Source: WDI

MIL _PERS
Military Personnel, total - MS.MIL.TOTL.P1
Source: WDI

UNEMPLOY
Unemployment, total as a percentage of the total labor force - SL.UEM.TOTL.ZS
Source: WDI

ROADS_NET
Roads, total network (km) - IS.ROD.TOTL.KM
Source: WDI

TAX_REV_P
Tax revenues as a percentage of GDP - GB.TAX.TOTL.GD.ZS
Source: WDI

TAX_REV
Tax revenue (current Local Currency Units) - GB.TAX.TOTL.CN
Source: WDI

MALE_SCHOOL
School enrollment, secondary, male as a percentage of gross enrollment - SE.SEC.ENRR.MA
Source: WDI

MALE_SCHOOL_NET
School enrollment, secondary, male as a percentage of net enrollment - SE.SEC.NENR.MA
Source: WDI

POP_DEN
Population density (People per square kilometer) - EN.POP.DNST
Source: WDI

POP_DEN_RUR
Population density rural (People per square kilometer) - EN.RUR.DNST
Source: WDI

AID_CAPITA
Aid received per capita (current US Dollar) - DT.ODA.ALLD.PC.ZS
Source: WDI

GINI
GINI index - SI.POV.GINI
Source: WDI

INCOME_4TH20
Income share held by the fourth 20th percentile of the population - SI.DST.04TH.20
Source: WDI

INCOME_1ST10
Income share held by the highest 10th percentile of the population - SI.DST.10TH.10

Source: WDI

INCOME_1ST20
Income share held by the highest 20th percentile of the population - SI.DST.05TH.20
Source: WDI

INCOME_10TH10
Income share held by the lowest 10th percentile of the population - SI.DST.FRST.10
Source: WDI

INCOME_5TH20
Income share held by the lowest 20th percentile of the population - SI.DST.FRST.20
Source: WDI

INCOME_2TH20
Income share held by the second 20th percentile of the population - SI.DST.02ND.20
Source: WDI

INCOME_3RD20
Income share held by the third 20th percentile of the population - SI.DST.03RD.20
Source: WDI

LAND_ARABLE
Land use, arable land as a percentage of land area - AG.LND.ARBL.ZS
Source: WDI

LAND_CROP
Land use, permanent cropland as a percentage of land area - AG.LND.CROP.ZS
Source: WDI

LAND_FOREST
Forest area as a percentage of land area - AG.LND.FRST.ZS
Source: WDI

LAND_CROP_IRRIG
Land use, irrigated land as a percentage of cropland - AG.LND.IRIG.ZS
Source: WDI

LAND_OTHER
Land use, other as a percentage of land area - AG.LND.OTHR.ZS
Source: WDI

VA_AGR
Agriculture, value added as a percentage of GDP - NV.AGR.TOTL.ZS
Source: WDI

VA_IND_MANF
Manufacturing, value added as a percentage of GDP - NV.IND.MANF.ZS
Source: WDI

VA_IND_TOT
Industry, value added as a percentage of GDP - NV.IND.TOTL.ZS
Source: WDI

VA_SERV
Services, etc., value added as a percentage of GDP - NV.SRV.TETC.ZS
Source: WDI

POP_0014_FEM
Population ages 0-14, female - SP.POP.0014.FE.IN
Source: WDI

POP_0014_MALE
Population ages 0-14, male - SP.POP.0014.MA.IN
Source: WDI

POP_0014_TOT
Population ages 0-14, total - SP.POP.0014.TO
Source: WDI

POP_0014_PTOT
Population ages 0-14 as a percentage of the total population - SP.POP.0014.TO.ZS
Source: WDI

POP_1564_FEM
Population ages 15-64, female - SP.POP.1564.FE.IN
Source: WDI

POP_1564_PTOT
Population ages 15-64 as a percentage of the total population - SP.POP.1564.IN.ZS
Source: WDI

POP_1564_MALE
Population ages 0-14, male - SP.POP.1564.MA.IN
Source: WDI

POP_1564_TOT
Population ages 0-14, total - SP.POP.1564.TO
Source: WDI

POP_65UP_FEM
Population ages 65 and above, female - SP.POP.65UP.FE.IN
Source: WDI

POP_65UP_MALE
Population ages 65 and above, male - SP.POP.65UP.MA.IN
Source: WDI

POP_65UP_FEM_PMALE
Population ages 65 and above, per 100 men - SP.POP.65UP.MF.ZS
Source: WDI

POP_65UP_TOT
Population ages 65 and above, total - SP.POP.65UP.TO
Source: WDI

POP_65UP_PTOT
Population ages 65 and above, percentage of the total population - SP.POP.65UP.TO.ZS
Source: WDI

POV_HEAD_NAT
Poverty headcount, national, as a percentage of the population - SI.POV.NAHC
Source: WDI

POV_HEAD_RUR
Poverty headcount, rural, as a percentage of the population - SI.POV.RUHC
Source: WDI

POV_HEAD_URB
Poverty headcount, urban, as a percentage of the population - SI.POV.URHC
Source: WDI

POP_RUR_PTOT
Rural population as a percentage of the total population - SP.RUR.TOTL.ZS
Source: WDI

POP_TOT
Population, total - SP.POP.TOTL
Source: WDI

POP_1524_MALE
Population ages 15-24
Source: WDI

PER_0014
Percentage of males ages 0-14 of the total population
Source: WDI

PER_1524
Percentage of males ages 15-24 of the total population
Source: WDI

C. FAO Stat and Derived

FAO_FOODAID
Food Aid All Donors - Wheat, Rice, Barley, Maize, Rye, Oats, Millet, etc.
Source: FAO Stat

D. Alesina et al. Fractionalization and Derived

ETHNIC
Ethnic Fractionalization.
Source: Alesina et al. (2003)

LANGUAGE
Linguistic Fractionalization.
Source: Alesina et al. (2003)

RELGION
Religious Fractionalization.
Source: Alesina et al. (2003)

4. Political Institutional Controls

A. Polity IV and Derived

DEMOC
Institutionalized Democracy Score. Ranges from 0, least democratic to 10, most democratic. In addition, the following variables are coded as such: Interruption Periods (-66), Interregnum Periods (-77), Transition Periods (-88).
Source: Polity IV

AUTOC
Institutionalized Autocracy Score. Ranges from 0, least autocratic to 10, most autocratic. In addition, the following variables are coded as such: Interruption Periods (-66), Interregnum Periods (-77), Transition Periods (-88).
Source: Polity IV

POLITY
Combined Polity Score (DEMOC - AUTOC). The difference between Polity IV's measure of democracy and its measure of autocracy. Values range from –10 to 10.
Source: Polity IV

POLITY2_IV
Revised Polity Score. Same as Polity2 variable above from Fearon and Laitin (2003). (DEMOC - AUTOC). The difference between Polity IV's measure of democracy and its measure of autocracy. Values range from –10 to 10.
The revised polity score fills in missing values based on the following coding: when polity = -66, set polity2 = NULL, when polity = -77, set polity2 = 0, when polity = -88, extrapolate based previous and subsequent values.
Source: Polity IV

DURABLE
Regime Durability.
Source: Polity IV

XRREG
Regulation of Chief Executive Recruitment
Source: Polity IV

XRCOMP
Competitiveness of Executive Recruitment
Source: Polity IV

XRCOMP
Competitiveness of Executive Recruitment
Source: Polity IV

XROPEN
Openness of Executive Recruitment
Source: Polity IV

XCONST
Executive Constraints (Decision Rules)
Source: Polity IV

PARREG
Regulation of Participation
Source: Polity IV

PARCOMP
The Competitiveness of Participation
Source: Polity IV

EXREC
Executive Recruitment Concept
Source: Polity IV

EXCONST
Executive Constraints Concept
Source: Polity IV

POLCOMP
Political Competition Concept
Source: Polity IV

B. Database of Political Institutions (DPI) and Derived

MILITARY
Is Chief Executive a military officer? Coded 1 if "Yes," 0 if "No," and Missing(.) if information not available.
Source: DPI

EXECRLC
Party of the Executive. Right (R); Left (L); Center (C); N/A (Missing(.))
Source: DPI

EXECNAT
Party of the Executive Nationalist? Coded 1 if "Yes," 0 if "No," and Missing(.) if information not available.
Source: DPI

EXECRURL
Party of the Executive Rural? Coded 1 if “Yes,” 0 if “No,” and Missing(.) if information not available.
Source: DPI

EXECREG
Party of the Executive Regional? Coded 1 if “Yes,” 0 if “No,” and Missing(.) if information not available.
Source: DPI

EXECREL
Party of the Executive Religious? Coded 1 if “Yes,” 0 if “No,” and Missing(.) if information not available.
Source: DPI

EXECAGE
Party of the Executive - Time Since Formation under this name?
Source: DPI

ALLHOUSE
Does party of executive control all of the relevant houses? Coded 1 if “Yes,” 0 if “No,” and Missing(.) if information not available.
Source: DPI

HERFGOV
Legislature Herfndahl Index Government
Source: DPI

HERFOPP
Legislature Herfndahl Index Opposition
Source: DPI

OPPFRAC
Legislature Opposition Fractionalization
Source: DPI

CHECKS
Checks and Balances.
Source: DPI

AUTON
Are there autonomous regions? Coded 1 if “Yes,” 0 if “No,” and Missing(.) if information not available.
Source: DPI

MUNI

Are the municipal governments locally elected?
"0 if neither local executive nor local legislature are locally elected. 1 if the executive is appointed, but the legislature elected. 2 if they are both locally elected."[6] Missing(.) if no information, or no evidence of municipal governments.
Source: DPI

STATE
Are the state / province governments locally elected?
"0 if neither local executive nor local legislature are locally elected. 1 if the executive is appointed, but the legislature elected. 2 if they are both locally elected."[7] Missing(.) if no information, or no evidence of state governments.
Source: DPI

AUTHOR
Do sub-national governments have extensive tax, spending, or regulatory authority? Coded 1 if "Yes," 0 if "No," and Missing(.) if information not available.
Source: DPI

STCONST
Are the constituencies of the senators the states / provinces? Coded 1 if "Yes," 0 if "No," and Missing(.) if information not available.
Source: DPI

C. Freedom House and Derived

FH_CIV
Civil Liberties Rating. Barro Transformation: (7 – FH Measure) / 6. The original Freedom House data ranges from 1, the highest level of freedom, to 7, the lowest level of freedom. After the Barro transformation, the data range from 1, the highest level of freedom, to 0, the lowest level of freedom.
Source: Freedom House

FH_POL
Political Rights Rating. Barro Transformation: (7 – FH Measure) / 6. The original Freedom House data ranges from 1, the highest level of freedom, to 7, the lowest level of freedom. After the Barro transformation, the data range from 1, the highest level of freedom, to 0, the lowest level of freedom.
Source: Freedom House

---

[6] Keefer, Philip. DPI2000 Database of Political Institutions: Changes and Variable Definition. Development Research Group, The World Bank. March 2002, p 21.
[7] Ibid, p21.

D. The Logic of Political Surival Data Set and Derived

S
Selectorate Size.
The selectorate is the broader group from which the winning coalition is drawn. It is constructed from the Polity variable Legislative Selection (LEGSELEC). Larger values of S represent larger selectorate sizes.
Source: The Logic of Political Survival Data Set

W
Winning Coalition Size.
W is a composite index based on data from Polity IV and Banks (1996). Specifically, W combines XRCOMP (the competitiveness of executive recruitment), XROPEN (the openness of executive recruitment) and PARCOMP (the competitiveness of participation) from Polity IV with REGTYPE from Banks (civilian character of regime). W takes on the following values: 0, 0.25, 0.5, 0.75, and 1.0. Larger values of W represent larger coalition sizes.
Source: The Logic of Political Survival Data Set

WoverS
W/S - Loyalty Norm
Source: The Logic of Political Survival Data Set

E. Barro (1991) and Derived

SOC
Socialist Country Dummy. Coded 1 if “Yes,” 0 Otherwise.
Source: Levine and Renelt (1992)

7. References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. (2003). "Fractionalization", *Journal of Economic Growth*, 8(2), 155-194.

Barro, Robert. (1991). "Economic Growth in a Cross Section of Countries", *Quarterly Journal of Economics*, 106(2), 407-443.

Beck, Thorsten, George Clarke, Alberto Groff, Philip Keefer, and Patrick Walsh. (2001). "New tools in comparative political economy: The Database of Political Institutions", *World Bank Economic Review*, 15: 1, 165-176.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. (2003). *The Logic of Political Survival*. Cambridge: MIT Press.

Fearon, James and David Laitin. (2003). "Ethnicity, Insurgency and Civil War." *American Political Science Review*, 97(1), 75-90.

Fearon, James and David Laitin. "Additional Tables for 'Ethnicity, Insurgency and Civil War.'" Stanford University, February 6, 2003.

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

Huffman, G.J., R.F. Adler, B. Rudolf, U. Schneider, P.R. Keehn (1995): Global Precipitation Estimates Based on a Technique for Combining Satellite-Based Estimates, Raingauge Analyses and NWP Model Information. Journal of Climate 8(5), p. 1284 - 1295.

Huffman, G.J., R.F. Adler, P.A. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, U. Schneider (1997): The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset. Bull. Americ. Meteor. Soc. 78(1), 5-20.

Huffman, G.J., R.F. Adler, M.M.. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B. McGavock, J. Susskind (2000/2001): Global Precipitation at One-Degree Daily Resolution From Multi-Satellite Observations. Accepted by J. Hydrometeorology.

Levine, Ross, and David Renelt. (1992). "Sensitivity Analysis of Cross-Country Growth Regressions", American Economic Review, LXXXII, 942-963.

Rudolf, Bruno. (2000). "Satellite-based Global Precipitation Estimates and Validation Results", Proceedings of the SAF Workshop on Climate Monitoring (http://www.eumetsat.de/en/area2/proceedings/eump31/pdf/rudolf.pdf).

Sambanis, Nicholas. (2001). “Do Ethnic and Non-Ethnic Civil Wars Have the Same Causes? A Theoretical and Empirical Inquiry (Part 1),” Journal of Conflict Resolution vol. 45, no. 3.

Sambanis, Nicholas. (2001a). “A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War”, unpublished manuscript, Yale University.

Xie, P. and P.A. Arkin, 1997: Global Precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull. Amer. Meteorol. Soc. 78, 2539-2558.

Addendum-

In our paper "Economic Shocks and Civil War: An Instrumental Variables Approach" (*Journal of Political Economy*, 2004) we examine the effects of economic growth shocks (as instrumented by rainfall shocks) on civil conflict. There are two ways of conceiving of civil conflict. One is a broad definition that aims to capture a country's civil war **involvement**. As per this broad definition countries where civil wars occur as well as countries that participate in other countries' civil conflicts are coded as a 1 on the civil war dummy variable. The second is a narrow definition of civil war that aims solely to capture the **location** of civil conflict. As per this definition a country is only coded as 1 if there is a civil war taking place within its boundaries. We note that the former broad definition is consistent with the causal story proposed in our paper. Our proposed causal mechanism is that adverse economic shocks (instrumented by rainfall shocks in our paper) make it easier to recruit fighters for civil conflicts. This effect of rainfall shocks could apply irrespective of whether fighters are being recruited for conflicts within their own countries or abroad.

The empirical difference between the two definitions is, however, close to moot in the African context because there are very few cases of interventions across borders: a shift from the first to the second definition (using the Uppsala/PRIO Armed Conflict Dataset version 1.2a, the dataset available when we wrote the original paper) requires recoding at most 25 out of 743 observations, or only about 3% of all cases.

Unsurprisingly, results hardly vary when we shift from the first to the second definition given the very small number of cases affected. The minor differences are as follows (see Table A below for the details):

1) In our paper we emphasize results using the threshold of at least 25 annual deaths to define a civil conflict. We do so because we think this comprehensive definition is more appropriate than the higher 1000 death threshold for smaller countries, and Africa (the focus of our analysis) abounds in small countries. This measure is best thought of as the most complete measure of armed civil conflicts, and is not a measure of "small conflicts" alone: it captures small and large conflicts. As the first four columns of Table A show, the significance level of our core results is almost entirely unaffected by a shift to the narrower definition, and the coefficient values change by very little, less than 5%. *It is fair to say that the main results of our paper are essentially unchanged when we shift the definition from the broad to the narrow definition of civil conflict involvement.*
2) As we showed in our original paper (Table 4, last column) the results for the higher 1000 death threshold are only borderline statistically significant at the 90% level. This can be readily explained by the fact that this is a more restrictive cut at the data, excluding many conflicts that should legitimately be counted in the African context (namely, those with annual death tolls between 25 and 1000). The effect on our results of shifting to a narrow definition of conflict, at any rate, is not substantial. The sign is unchanged, the p-value falls only slightly from 0.08

to 0.12 and the coefficient value drops by less than a third, though remains high. The fact that results using the annual 1000 death armed conflict threshold are weaker than for the comprehensive 25 death threshold was already clearly communicated in appendix Table C3 of our original paper via a series of robustness checks on multiple alternative data sets that use the annual 1000 death threshold, so there is little new information conveyed by the recoding discussed above.

In sum, the minor recoding of cases demanded by a shift to the narrow definition of armed conflict has little impact on the empirical results in our 2004 *JPE* paper. As such we interpret the analysis with the narrow definition of armed conflict as yet another valuable robustness check that confirms the main findings in the paper. For analysts who would like to use the narrow coding, we have updated the dataset on our websites. The new variables and their definitions are as follows:

LOCATION: Location of Conflict. 0=Country is not listed as the location of conflict; 1=Country is listed as the location of a Minor Conflict; 2=Country is listed as the location of an Intermediate Conflict; 3=Country is listed as the location of a War

ANY_PRIO_NAR: Dichotomous variable with death threshold of 25. Coded 1 if (TYPE3 equals 1, 2, or 3) or (TYPE4 equals 1, 2, or 3 & LOCATION is greater than 0).

WAR_PRIO_NAR: Dichotomous variable with death threshold of 1000. Coded 1 if (TYPE3 equals 3) or (TYPE4 equals 3 & TYPE3 equals 0 & LOCATION is greater than 0).

We are very grateful to Kristian Skrede Gleditsch (University of Essex) and Peter Sandholt Jensen (University of Aarhus) for suggesting these recodings, and for useful and constructive discussions about the empirical results.

Edward Miguel (University of California, Berkeley)
Shanker Satyanath (New York University)
Ernest Sergenti (New York University)

February 2007

Table A: Economic Growth and Civil Conflict – Broad and Narrow Codings

	Dependent variable: Civil conflict ≥ 25 deaths				Civil conflict ≥ 1000 deaths	
Explanatory variable	Broad (1)	Narrow (2)	Broad (3)	Narrow (4)	Broad (5)	Narrow (6)
Economic growth rate, t	-.41	-.50	-1.13	-.91	-1.48*	-1.04
	(1.48)	(1.26)	(1.40)	(1.21)	(.82)	(.65)
	[.782]	[.694]	[.425]	[.454]	[.080]	[.116]
Economic growth rate, t-1	-2.25**	-2.20**	-2.55**	-2.45**	-.77	-.41
	(1.07)	(1.05)	(1.10)	(1.11)	(.70)	(.62)
	[.043]	[.042]	[.026]	[.033]	[.278]	[.519]
Log(GDP per capita), 1979	.053	.069				
	(.098)	(.097)				
	[.592]	[.482]				
Democracy (Polity IV), t-1	.004	.007				
	(.006)	(.006)				
	[.484]	[.226]				
Ethno-linguistic fractionalization	.51	.43				
	(.39)	(.38)				
	[.203]	[.262]				
Religious fractionalization	.22	.04				
	(.44)	(.42)				
	[.620]	[.917]				
Oil exporting country	-.10	-.07				
	(.22)	(.22)				
	[.637]	[.738]				
Log(mountainous)	.060	.080				
	(.058)	(.058)				
	[.308]	[.173]				
Log (national population), t-1	.159*	.128				
	(.093)	(.088)				
	[.093]	[.156]				
Country fixed effects	No	No	Yes	Yes	Yes	Yes
Country-specific time trends	Yes	Yes	Yes	Yes	Yes	Yes
Root MSE	0.36	0.33	0.32	0.29	0.24	0.20
Number of observations	743	743	743	743	743	743

Table A Notes: Results presented in columns (1), (3), and (5) are identical to those presented in Miguel, Satyanath, and Sergenti (2004), Table 4, columns (5), (6), and (7) respectively and are from the sample using the broad definition of conflict. Results presented in columns (2), (4), and (6) are from the sample using the narrow definition of conflict. Huber robust standard errors are in parentheses. P-values are in brackets. Regression disturbance terms are clustered at the country level. The instrumental variables for economic growth in all regressions are growth in rainfall, t and growth in rainfall, t-1. A country-specific year time trend is included in all specifications (coefficient estimates not reported).

* Significantly different from zero at 90 percent confidence.

** Significantly different from zero at 95 percent confidence.

*** Significantly different from zero at 99 percent confidence.