Threats to Leaders' Political Survival and Pro-Government Militia Formation

Konstantin Ash

Replication Data Codebook

Instructions for Creating Tables and Figures

The STATA .do file ii_rebelterror_main.do can replicate each table in the main article and the appendix. Figure 2 in the main article, along the spatially lagged dependent variable included in a model in the appendix (Table 13) and Figure 1 in the appendix can be replicated using instructions embedded in the R-Code. To replicate Figure 1 in the main paper, ArcGIS is necessary to open the .mxd file in the folder Figure1Replication. The shapefiles and rasters are also included in the folder, already overlapped and can be adjusted to create the three images seen in Figure 1.

Sourcing and Documentation for Variables in Data

The following section provides sources for every variable included in the main replication file: ii_rebelterror_main_labels.dta. When possible, the labels are used as sub-headings.

ISO Country Code, Country Name

The first two variables contain the 3-letter ISO country code for the country of the observation and the name of the country, in accordance with the Global Terrorism Database's (GTD) coding.

Region Name (GADM), Region Name

There are two variables for the name of the region in the observation. One, with special characters that STATA does not display from the Global Administrative Areas Shapefile Dataset (GADM) (http://www.gadm.org/. The other, the identical names with the special characters replaced.

Group Name

The name of the insurgent group or grouping that corresponds to the observation. Names are taken or aggregated from the GTD.

Year

Year of the observation.

Average Regional Nighttime Lights

Calculated using nighttime light data from the National Oceanic and Atmospheric Association's (NOAA) Defense Satellite Meteorological Program (DSMP) by taking the average nighttime lights

in a region, based on shapefiles obtained from GADM. The average is obtained using zonal statistics in ArcGIS.

Average National Nighttime Lights

This measure is calculated in the same way as the above measure, only using national, not regional, shapefiles to derive nighttime light statistics.

Regional Population Density

Raster files that encode population density to geographic space for five year intervals between 1990 and 2005 are obtained from Columbia University's Socioeconomic Data and Applications Center (SEDAC). The Center geocodes population density figures allowing for calculation of regional averages using zonal statistics and border shapefiles. This variable is the regional average (regional shapefiles obtained from SEDAC.

National Population Density

This is the corresponding national population density variable. It is calculated in the same way as the regional variable only using national border shapefiles.

Urban Population as a percent of total regional population

The above center also has geogreferenced measures of population counts. These averages are taken using both regional shapefiles and regional shapefiles that only contain urban space in a region. Borders for urban areas around the world are also available from SEDAC (for International Earth Science Information Network at Columbia University et al., 2011). The urban population count (from the urban regional shapefiles) is divided by the total regional population to obtain this measure.

Regional gross cell product (G-Econ)

The G-Econ data contains estimates for the total gross product of 50km by 50km cells of all landed area in the world (Nordhaus et al., 2006). This can be aggregated into borders to create average regional gross cell products using the mentioned GADM shapefiles. The variable contains data in five year intervals from 1990 to 2005.

National gross cell product (G-Econ)

G-Econ data was averaged using national shapefiles from GADM to get this measure.

Mean Elevation

The G-Topo (Survey, 2006) contains data for all elevation data of landed area around the world at a 30 meter resolution. Using GADM borders it was possible to make zonal averages for all of the

regions included in the study. The variable does not vary over time.

Mean Capital Distance

The PRIO-Grid dataset (Tollefsen et al., 2012) contains population density data for each point of landed area a 1 sq. km. resolution. This was averaged into zonal statistics using GADM data on regions. The variable is time invariant.

Distance to Petroleum Deposits

Data on geographic location of petroleum deposits is coded dichotomously based on shapefiles from the PETRODATA (Lujala et al., 2007) data-set on oil deposits. The variable is coded as one if there is an overlapping petroleum deposit in a region and zero otherwise. The variable is time invariant.

Distance to Diamonds

Similar to petroleum deposits, a shapefile of global diamond deposits yielded the location of diamond mines (Lujala et al., 2005). This allowed for a dichotomous variable of whether a region contained a diamond mine or not.

Number of terrorist attacks (military and police targets excluded)

Taken from a collapsed measure of each terrorist attack that occurred in a given country attributable to a particular group in a given year, distributed based on region. The count measure excludes 1993 as it is missing from the GTD data-set. Attacks listed by the GTD as against military or politce targets are also excluded.

Number of insurgent attacks (military targets only)

Taken from a collapsed measure of each terrorist attack that occurred in a given country attributable to a particular group in a given year, targeting a national military, distributed based on region. The count measure excludes 1993 as it is missing from the GTD data-set.

Regional Light Density

Taking the average regional nighttime lights and dividing them by population density. Specifically, the average is divided by the square root of population density obtained from SEDAC (University Center for International Earth Science Information Network, 2011) using the same zonal statistics method. The result is a measure of average per capita nighttime lights in a given region.

National Light Density

Identical calculations to previous measure only divided by national averages.

Regional GCP Density, National GCP Density

These average density measures are calculated in the same way as the light density measures, only using average gross cell product for a region or country.

Interregional Light Inequality

The regional light density divided by the national light density produces a measure of relative inequality in nighttime light provision for a given region in a given year.

Interregional Wealth Inequality

The regional GCP density divided by the national GCP density produces a measure of relative inequality in GCP for a given region in a given year.

Logs of Interregional Wealth and Light Deviations

Natural logs of the previous two measures.

National Population Count

Obtained from SEDAC shapefiles, zonal statistics for the sum of the number of people within the borders of a given country. Available at five year intervals from 1990 to 2010. Data carried forward for missing years.

Regional Population Count

Obtained from SEDAC shapefiles, zonal statistics for the sum of the number of people within the borders of a given region. Available at five year intervals from 1990 to 2010. Data carried forward for missing years.

Regional Population as pct. of National Population

Regional population count divided by national population count.

activemax, active

Classification of whether group is an active insurgency in a given year based on the number and types of attacks committed. Specific instructions for calculating each of these are at the beginning of the .do file.

Capital City

Author-coded dichotomous variable of whether a region contains a capital city. Time invariant.

Log of Interregional Light Deviation Squared, Log of Regional Wealth Deviation Squared

Both the interregional wealth and light inequality measures were squared and then their natural log was taken to examine non-linear effect.

Low Light Deviation, High Light Deviation

This technique, in line with Stewart et al. (2008) examines only variation on one side of the distribution of deviation, either high or low, intended to examine non-linear effects. More detail on the formula is provided in the appendix.

Region Borders Insurgency

Dichotomous variable as to whether a region bordered an location whether there was an active insurgency (attacks against the military according to the GTD). The variable was author-coded and time-variant.

Number of insurgent attacks (military targets only) (with 1993)

An imputed version of the mentioned count measure of the number of attacks against military targets by a group in a given region-year. The imputation is done to fill 1993, which is missing from the GTD.

Number of terrorist attacks (military and police targets excluded) (with 1993)

An imputed version of the mentioned count measure of the number of terrorist attacks, except against police and military targets by a group in a given region-year. The imputation is done to fill 1993, which is missing from the GTD. This is the dependent variable for the study.

Interaction between Log of Interregional Light Deviation and bordering Insurgency

The main independent variable multiplied by a dichotomous variable of whether a region borders a region experiencing insurgency.

Active Insurgency

A dichotomous measure based on whether a region experiences attacks by a particular group against military targets in a given year. If so, coded as one, if not, zero. Time variant.

Interaction between Log of Interregional Light Deviation and Active Insurgency

The main independent variable multiplied by a dichotomous variable of whether a region is experiencing insurgency in a given year.

Per Capita Gross Regional Product

Gross regional product data was available for Russia, Turkey, the Philippines, India and Colombia. The measures are generally taken as direct per capita gross cell product data from national statistical offices. These offices are listed on page 10 of the appendix.

Per Capita Gross Domestic Product

Corresponding per capita GDP measure from national statistical offices of Russia, Turkey, the Philippines, India and Colombia. Sources are same as the above GRP measure.

Interregional Wealth Deviation (national statistical offices)

The measure of gross regional product of Russia, Turkey, the Philippines, India and Colombia divided by gross national product for a given year.

Log of Interregional Wealth Deviation (national statistical offices)

Natural log of the previous wealth deviation measure.

exposure, insexposure

Two exposure variables - logs of the maximum possible attacks committed by a group in a given year, coded as the same value across each group-year. The variables are derived by taking the

cumulative amount of attacks committed by a group in a given year, or the cumulative number of insurgent attacks and then taking the natural log.

Total Attacks by a Group in a Region, Total Attacks by another Group in a Region

The sum of all attacks, across time, that a group committed in a region of a country and the sum of all attacks that any other groups committed in the same region of the country in a given year. The latter measure is intended to measure inter-rebel competition.

Diamond Deposits

Dichotomous measure of whether a given region contained diamond deposits. Coded based on shapefile data from Lujala et al. (2007).

Petroleum Deposits

Dichotomous measure of whether a given region contained petroleum deposits. Coded based on shapefile data from Lujala et al. (2005).

Interaction between Log of Interregional Wealth Deviation and bordering Insurgency

The log of interregional wealth deviation multiplied by a dichotomous variable of whether a region bordered one with attacks against military targets in that year.

Interaction between Log of Interregional Wealth Deviation and Active Insurgency

The log of interregional wealth deviation multiplied by a dichotomous variable of whether a region had attacks against military targets in that year.

Rebels' Identity Group in Region, identitygroup

Dichotomous variable of whether the region contained substantial geographical space that was occupied by members of the same identity. Data taken from shapefiles from the GeoEPR (Wucherpfennig et al., 2011) and the Gulf/2000 project (Izady, 2013).

Square Root of Population Density

Measure of the square root of each region's population density in a given year - only used for demonstration, not in any models.

attacks_c, Terrorist Attacks against Government Targets (imputed)

Aggregated count of the amount of terrorist attacks that target what the GTD defines as a government target. An additional measure, imputing missingness from 1993 is also included.

attacks_c, Terrorist Attacks against Civilian Targets (imputed)

Aggregated count of the amount of terrorist attacks that target what the GTD defines as a civilian target. An additional measure, imputing missingness from 1993 is also included.

grouping, foreign_support

'grouping' is a dichotomous variable of whether a certain group was actually a amalgamation of several group due to imprecise group naming by the GTD. Foreign support is, as defined by Salehyan et al. (2011), any support a group received from foreign actors. Both measures are dichotomous and time invariant.

active_count

A variable of how many consecutive years a group has committed insurgent attacks. Derived from the "Active insurgency" variable and intended to capture variation in group behavior within spells of insurgent violence.

excluded, excluded_insurgency, excluded_border

Data from the GeoEPR on whether a region contains an ethnic group that has been deemed by the EPR as either "powerless" or "discriminated." Two subsequent measures provide interaction variables of the measure with active insurgency and bordering an active insurgency.

included

Data from the GeoEPR on whether a region contains an ethnic group that has been deemed by the EPR as a "senior partner," "dominant" or "monopoly." Dichotomous, time variant.

lightineq_orig

The original measure of Interregional Light Inequality without square roots of density (just divided by unaltered population density). Intended to show broad similarity of the two measures.

Longitude, Latitude

The center latitude and longitude of every region in the data-set. Derived from shapefiles included in replication data through R.

References

for International Earth Science Information Network at Columbia University, Center, International Food Policy Research Institute IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical CIAT (2011). *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extents Grid*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Izady, Michael (2013). The gulf/2000 project. Maps and Statistical Collections: http://gulf2000.columbia.edu/maps.shtml.

Lujala, Päivi, Jan Ketil Rød, and Nadja Thieme (2007). Fighting over oil: Introducing a new dataset. *Conflict Management and Peace Science 24*(3), 239–256.

Lujala, PÃd'ivi, Nils Petter Gleditsch, and Elisabeth Gilmore (2005). A diamond curse?: Civil war and a lootable resource. *Journal of Conflict Resolution 49*(4), 538–562.

Nordhaus, William, Qazi Azam, David Coderi, Kyle Hood, Nadejda Makarova Victor, Mukhtar Mohammed, Alexandra Miltner, and Jyldyz Weiss (2006). The g-econ database on gridded output: Methods and data. Geographically-based Economic Data.

Salehyan, Idean, Kristian Skrede Gleditsch, and David E. Cunningham (2011, 10). Explaining external support for insurgent groups. *International Organization 65*, 709–744.

Stewart, Frances, Graham K. Brown, and Arnim Langer (2008). Major findings and conclusions on the relationship between horizontal inequalities and conflict. In F. Stewart (Ed.), *Horizontal Inequalities and Conflict: Understanding Group Violence in Multiethnic Societies*, pp. 285–300. Houndmills, UK: Palgrave MacMillan.

Survey, U.S. Geological (2006). Global digital elevation model (gtopo30). EROS Data Center Distributed Active Archive Center.

Tollefsen, Andreas Forø, Håvard Strand, and Halvard Buhaug (2012). Prio-grid: A unified spatial data structure. *Journal of Peace Research 49*(2), pp. 363–374.

University Center for International Earth Science Information Network (2011). *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Density Grid*. At Columbia, and International Food Policy Research Institute IFPRI and The World Bank and Centro Internacional de Agricultura Tropical CIAT. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Wucherpfennig, Julian, Nils B. Weidmann, Luc Girardin, Lars-Erik Cederman, and Andreas Wimmer (2011). Politically relevant ethnic groups across space and time: Introducing the geoepr dataset. *Conflict Management and Peace Science*.