

## **Pierre Englebert**

### ***Africa: Unity, Sovereignty, and Sorrow***

#### **Methods, Measurements, and Sources for Data Used in Chapter Two**

Chapter Two uses two different data sets. The first one (used for Table 2.3) is longitudinal. It uses countries as units of analysis, and covers the 1960-1999 period in five-year intervals (1960-1964, 1965-1969, etc.). It has a maximum of 1,219 observations, 932 of which are included in the model in Table 2.3. The second one (used for Table 2.4) is cross-sectional. It uses “minorities at risk” (MAR) as units and measures variables over the 1990-2003 period.<sup>1</sup> This dataset identifies 338 “politically-active communal groups in all countries with a current population of at least 500,000.” In Table 2.4, respectively 310 and 288 observations had enough data to be included in the model.

Both data sets are used to analyze the determinants of separatist conflict and separatist sentiment, the two dependent variables in Tables 2.3 and 2.4. The measurement of separatist conflict is adapted from the Peace Research Institute of Oslo’s (PRIO) measure of “domestic conflicts based on territorial incompatibility,” which refers to struggles over separatism and autonomy.<sup>2</sup> PRIO provides the years during which such conflicts take place together with a measure of their intensity in terms of battle deaths. When a conflict has at least 25 battle deaths in a year and no more than 1,000 overall, it is classified as “minor.” An “intermediate” conflict has between 25 and 1,000 battle deaths per year with an accumulated total of more than 1,000. Finally, “wars” have at least 1,000 battle-related deaths per year. In the country dataset, I give these three levels of conflict values of 1, 2, and 3, and I enter the highest applicable value (including 0) for each five-year period.<sup>3</sup> I chose this measure of intensity of conflict over the more common measure of onset because separatist efforts in Africa appear not only fewer but also more lukewarm than elsewhere.<sup>4</sup> Because of the relative scarcity of separatist conflicts across space and time, this variable is oddly distributed with 1,099 observations scoring 0 and 120 positive ones. The consequent skewness contributes to the overall predictive weakness of the model, making it hard to establish the validity of the underlying theories and to ascertain the specificity of Africa.

One way around this problem is to consider secessions a statistical “rare event” and adjust the data set in ways that correct for this scarcity by selecting a sample of observations that includes all the positive ones and a random selection of those scoring 0 of about twice the size of the positive ones, resulting in a sample of about 330 observations. This is a solution proposed by Gary King and Langche Zeng.<sup>5</sup> Doing so increases the variance of the dependent variable, while all information on actual

---

<sup>1</sup> For details and data, see <http://www.cidcm.umd.edu/inscr/mar/>.

<sup>2</sup> Gleditsch et al., “Armed Conflict...”

<sup>3</sup> Using a dummy version of the variable, scoring 1 in any period of separatist activity, did not yield significantly different findings.

<sup>4</sup> For works using the onset of conflict as the dependent variable, see Sambanis, “Partition...,” and Collier and Hoeffler, “The Political Economy...”

<sup>5</sup> “Logistic Regression in Rare Events Data.” Department of Government, Harvard University, 1999 (<http://gking.harvard.edu>); and Gary King and Langche Zeng, “Explaining Rare Events in International Relations,” *International Organization*, Summer 2001, 55(3):693-715.

separatist instances is included together with a representative sample of the non-separatist cases. I tried this method. It yielded similar results to those of the larger model.

In the minority dataset, I coded each group as to whether and how intensely it engaged in a separatist conflict over the period 1990-2003, along the 0-3 scale developed above from the PRIO data. Forty-eight groups qualified as separatist with at least minor violence since 1990.<sup>6</sup>

The measurement of separatist sentiment or preferences is adapted from the MAR data, which includes such a variable based on a qualitative study of each group by MAR researchers. This variable does not require any threshold of violence to score a positive value. It takes on values ranging from 0 (no separatist sentiment) to 3 (active separatist or autonomy movement), with 1 and 2 respectively indicating latent and historical separatist sentiments. For the country dataset, I calculated the sum of the separatist scores of all identified minorities. For example, Afghanistan has four such minorities. The Hazaras score 1 on the MAR separatism index and the three other ones (Pashtuns, Tajiks and Uzbeks) score 0. Consequently, Afghanistan as a whole receives a score of 1. In Angola, the Bakongo score 2, Cabinda 3 and the Ovimbundu 0. Angola then receives a score of 5. Because the MAR separatism data is only valid for the 1990s, I only use data from the 1990-1995 period with this dependent variable. In other words, this leads me to abandon the time dimension of the dataset and focus simply on a cross-national test of the theories. As a result there are only 143 observations in the “sentiment” model of Table 2.3.

### Country Dataset

Model (1) in Table 2.3 is a Prais-Winsten regression, with heteroskedastic panels-corrected standard errors, 1<sup>st</sup>-order autocorrelation. The cross-sectional variable is “cow,” the country numeric code from the Correlates of War data set. The time variable is year (5-year increments).

The STATA command was:

```
xtpcse terrint africa lac gdp language allfailure priogovt transition
postcoldwar mar pop, correlation(pсар1) hetonly
```

The STATA results are:

```
Group variable:   cow                Number of obs   =      932
Time variable:   year                Number of groups =      153
Panels:          heteroskedastic (unbalanced)  Obs per group: min =      1
Autocorrelation: panel-specific AR(1)          avg = 6.091503
                                                max =      8
Estimated covariances =      153          R-squared       = 0.1812
Estimated autocorrelations =      153      Wald chi2(10)   = 134.61
Estimated coefficients =      11          Prob > chi2     = 0.0000
```

-----						
	Het-corrected					
terrint	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
africa	-.4875161	.0765496	-6.37	0.000	-.6375504	-.3374817

<sup>6</sup> Using a simple separatist-conflict dummy as dependent variable yielded similar results.

lac	-.4384203	.0526746	-8.32	0.000	-.5416607	-.3351799
gdp	-4.37e-06	3.03e-06	-1.45	0.148	-.0000103	1.56e-06
language	.4084159	.0942982	4.33	0.000	.2235949	.5932369
allfailure	-.2003075	.0493876	-4.06	0.000	-.2971055	-.1035096
priogovt	.0742196	.0226956	3.27	0.001	.0297371	.1187022
transition	-.0482032	.017534	-2.75	0.006	-.0825691	-.0138372
postcoldwar	.1219399	.049824	2.45	0.014	.0242867	.2195932
mar	.3743858	.1137078	3.29	0.001	.1515226	.597249
pop	5.39e-07	2.57e-07	2.10	0.036	3.60e-08	1.04e-06
_cons	.185841	.0519636	3.58	0.000	.0839942	.2876879
-----						
rhos =	0	0	0	0	0 ...	0
-----						

For model (2), the command and results are as follows:

```
regress sumsep africa lac gdp language plural allfailure mar size if year == 1990, r
```

```
Linear regression                               Number of obs =    143
                                                F( 8, 134) =    4.53
                                                Prob > F      =    0.0001
                                                R-squared    =    0.4773
                                                Root MSE    =    4.169
```

-----						
sumsep	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
africa	-4.054823	1.234053	-3.29	0.001	-6.495566	-1.614081
lac	-4.119175	.9389606	-4.39	0.000	-5.976276	-2.262075
gdp	-5.96e-07	.0000466	-0.01	0.990	-.0000927	.0000916
language	3.411622	1.672001	2.04	0.043	.1046957	6.718548
plural	-1.104198	1.810672	-0.61	0.543	-4.685392	2.476996
allfailure	-1.858119	.6690178	-2.78	0.006	-3.18132	-.5349188
mar	2.271912	1.564175	1.45	0.149	-.8217539	5.365579
size	3.14e-06	1.11e-06	2.82	0.006	9.38e-07	5.34e-06
_cons	1.890367	1.679517	1.13	0.262	-1.431425	5.212159
-----						

The adjusted R-square is 0.4461.

### **List of variables used in the Country Dataset**

#### **terrint (dependent variable, model 1)**

Highest level of intensity of territorial conflict in a given period (1=minor, 2=intermediate, 3=war). Source: PRIO.

#### **sumsep (dependent variable, model 2)**

Sum for each country of the separatism score of its minorities, as assessed by Minorities at Risk's separatism index. Source: <http://www.cidcm.umd.edu/inscr/mar/>.

#### **africa**

Dummy variable. Takes the value of 1 for African countries; 0 otherwise.

**lac**

Dummy variable. Takes the value of 1 for Latin American and Caribbean countries; 0 otherwise.

**cow**

3-digit country code from Correlates of War.

**year**

first year of five-year period (e.g., 1995 for 1995-99)

**gdp**

Per capita GDP in thousands of constant 1995 dollars. Source: “rpcgdpdol” (WDI) or its predicted value inferred from “rgdpch” from Penn World Table (PWT) 6.1 (see table for conversion factors). Rpcgdpdol is in constant 1995 dollars. PWT 6.1 data is in 1996 prices. GDP is therefore in 1995 real prices or estimated 1995 prices (as predicted by 1996 prices). For Iraq (1998-2001) data from Economist. Iraq 1996-97 inferred from Economist data and growth rates for 1997 and 1996 given by “CountryWatch”. For Yugoslavia, [http://migration.ucdavis.edu/mn/archive\\_mn/jun\\_1999-09mn.html for years 1989-1997](http://migration.ucdavis.edu/mn/archive_mn/jun_1999-09mn.html_for_years_1989-1997). Says \$3,000 in 1989. Decreased by increments of \$300 a year to match about \$1,200 in 1995. Other data on Iraq, Myanmar (latest years), North Korea, Somalia, Liberia, Micronesia, and Marshall Islands derived and adjusted from miscellaneous *Country Reports* of the Economist Intelligence Unit. EIU yields \$76 for Somalia for 1990 in current terms. All of Somalia 1989 and after was entered as \$200. last available “real” entry, for 1988, was \$237. Chile is wrong in WDI. Corrected by using inferred value from RGDPCHE. For Bosnia-Herzegovina, 1995 value is used for 1992-4. For Senegal 2001, a growth rate of 2.71% was applied to 2000 figure (5.71% per CIA – 3% estimated population growth).

Conversion ratios for the imputation of missing values in GDP. For the 943 observations which have values for RGDPCHE but not for rpcgdpdol, the following ratios were used to predict the value of rpcgdpdol (this was used instead of regressing one on the other, which generates negative values). These predicted values were then added as observations in the variable GDP which is equal to rpcgdpdol if available, or the above-mentioned predicted value, or some alternative ad hoc source for a few missing cases as indicated above. For example, for all values of RGDPCHE below \$500 which had missing values for rpcgdpdol, the predicted value of GDP is RGDPCHE\*(187/433), etc.

```
. su rgdpch rpcgdpdol if rgdpch<500
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rgdpch	60	433.0162	54.84278	281.26	499.66
rpcgdpdol	30	187.5043	111.3222	84.72279	590.4184

```
. su rgdpch rpcgdpdol if rgdpch>500 & rgdpch<1000
```

Variable	Obs	Mean	Std. Dev.	Min	Max
----------	-----	------	-----------	-----	-----

```

    rgdpch |          674    781.3714    140.9254    500.73    999.53
    rpcgdpdol |          519    236.3297    103.2827    76.62579    686.3513

. su rgdpch rpcgdpdol if rgdpch>1000 & rgdpch<2000

    Variable |          Obs          Mean    Std. Dev.          Min          Max
-----+-----
    rgdpch |          1004    1410.482    286.8623    1000.02    1998.11
    rpcgdpdol |           844    425.3947    177.26    148.5691    1547.079

. su rgdpch rpcgdpdol if rgdpch>2000 & rgdpch<5000

    Variable |          Obs          Mean    Std. Dev.          Min          Max
-----+-----
    rgdpch |          1582    3306.696    865.0032    2003.64    4998.05
    rpcgdpdol |         1266    1343.278    713.841    177.7922    8398.532

. su rgdpch rpcgdpdol if rgdpch>5000 & rgdpch<10000

    Variable |          Obs          Mean    Std. Dev.          Min          Max
-----+-----
    rgdpch |          1081    7198.738    1437.578    5000.09    9993.23
    rpcgdpdol |           873    4564.669    2716.902    173.4685    16969.19

. su rgdpch rpcgdpdol if rgdpch>10000 & rgdpch<20000

    Variable |          Obs          Mean    Std. Dev.          Min          Max
-----+-----
    rgdpch |           809    14182.46    2756.834    10006.15    19997.28
    rpcgdpdol |           748    15723.45    6295.024    3081.983    34228.14

. su rgdpch rpcgdpdol if rgdpch>20000 & rgdpch<.

    Variable |          Obs          Mean    Std. Dev.          Min          Max
-----+-----
    rgdpch |           236    23209.7    3598.207    20000.82    43989.44
    rpcgdpdol |           236    30413.66    8036.664    15131.76    56372

```

### **maleseced2**

Male secondary school gross enrollment ratios, extrapolated from “maleseced”. Source: WDI.

### **oil**

Dummy variable for oil-exporting countries.

### **language**

Linguistic fractionalization index. Source: Alesina et al., 2002.

### **plural**

Size of largest ethno-linguistic group in a country, in decimals. Source: Fearon and Laitin.

### **allfailure**

Index of state failure. This variable is actually a measure of state strength. The higher the number, the stronger the state. Thus, the sign of the coefficient was inverted in the

tables to more intuitively capture the effects of failure itself. The variable “failure” is first generated by creating the principal component of the variables “law and order”, “bureaucratic quality” and “government stability” from the PRS data set. “failure” is then regressed on the absolute value of the “polity” variable from the Polity IV dataset. “allfailure” is either the value for “failure” or, if it is missing, the predicted value from the regression on polity. The coefficient of this variable is likely inflated by endogeneity of failure to separatist conflict.

**priogovt**

Non-separatist political violence; measures the number of years of non-separatist political violence (such as violence over the control of central government) in a country over each five-year period. Source: PRIO.

**polity2**

Index of the democratic nature of a regime, from authoritarian (-10) to democratic (10). Source: Polity IV dataset.

**transition**

To measure regime transitions, I created a dummy variable which takes on a value of 1 when a country’s political system is ranked as between -3 and 3 on the “Polity 2” variable from the Polity IV data set.<sup>7</sup> The “Polity 2” variable ranges from -10 (authoritarian regimes) to 10 (democracies). Values close to 0 on the “Polity 2” variable capture regimes which are neither authoritarian nor democratic. Although these would include regimes in transition, they can also be stable regimes which have characteristics of both authoritarianism and democracy. As a result, although it is unclear why these regimes would reduce the likelihood of separatism, this may not necessarily be a repudiation of the underlying hypothesis.

**postcoldwar**

Dummy variable, taking value of 1 for all years after 1989.

**mar**

% of a country’s population listed as “minority at risk” by the Minorities at Risk project. Used as a measure of repressive nature of government. Source: MAR.

**age**

Country age: year of data – year of independence.

**pop**

in millions

**size**

country size, in million square miles.

**split**

---

<sup>7</sup> For data and explanations, see <http://www.cidcm.umd.edu/inscr/polity/>.

dummy variable for territorial discontinuity.

### Group Dataset

Both models (1) and (2) in Table 2.4 are OLS regressions with heteroskedasticity-consistent standard errors. The data covers the period 1990 to 2003. Some variables are averaged or summed over the period; others are measured at the beginning of the period and labeled with the suffix "90." For model (1), the STATA command and results were:

```
regress terrint africa lac gdp90 maleseced90 oilandcoal groupcon belief
language discrimination90 allfailure90 autlost autonregion size ,r
```

```
Linear regression                               Number of obs =      310
                                                F( 13, 296) =      3.97
                                                Prob > F          =    0.0000
                                                R-squared         =    0.2546
                                                Root MSE         =    .68625
```

terrint	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
africa	-.3910483	.1556544	-2.51	0.013	-.6973778	-.0847189
lac	-.4078542	.1068021	-3.82	0.000	-.618042	-.1976665
gdp90	3.52e-06	5.63e-06	0.63	0.532	-7.56e-06	.0000146
maleseced90	-.0012073	.0020065	-0.60	0.548	-.0051561	.0027416
oilandcoal	.0928611	.1303777	0.71	0.477	-.1637237	.3494459
groupcon	.0392029	.0320826	1.22	0.223	-.023936	.1023418
belief	.0321348	.0278794	1.15	0.250	-.0227321	.0870017
language	.2690201	.1939822	1.39	0.167	-.112739	.6507791
discrimin~90	.0982517	.034007	2.89	0.004	.0313256	.1651777
allfailure90	.1199302	.0412809	2.91	0.004	.0386891	.2011714
autlost	.1427434	.0493017	2.90	0.004	.0457171	.2397698
autonregion	.3360731	.1322295	2.54	0.012	.0758441	.5963022
size	-3.41e-08	2.01e-08	-1.70	0.090	-7.36e-08	5.38e-09
_cons	-.1755587	.1904661	-0.92	0.357	-.5503981	.1992807

For model (2), the STATA command and results were:

```
regress separatism africa lac maleseced90 oilandcoal ecdiffabs regional grou
> pcon language diaspora allfailure90 autlost autonregion bordering pop,r
```

```
Linear regression                               Number of obs =      288
                                                F( 14, 273) =     33.55
                                                Prob > F          =    0.0000
                                                R-squared         =    0.4378
                                                Root MSE         =    .99947
```

separatism	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
africa	-.6918356	.2097403	-3.30	0.001	-1.10475	-.2789216
lac	-1.05136	.1979376	-5.31	0.000	-1.441038	-.661682
maleseced90	-.000824	.0034596	-0.24	0.812	-.0076348	.0059868
oilandcoal	.0727389	.1926052	0.38	0.706	-.3064413	.4519191
ecdifffabs	-.0369287	.0417348	-0.88	0.377	-.1190916	.0452343

regional	.6103697	.1784189	3.42	0.001	.2591179	.9616215
groupcon	.0684163	.074272	0.92	0.358	-.0778023	.2146349
language	.0720689	.2764235	0.26	0.795	-.4721238	.6162616
diaspora	.1032542	.145864	0.71	0.480	-.1839071	.3904155
allfailure90	.1431242	.0602133	2.38	0.018	.0245827	.2616657
autlost	.4311234	.0717548	6.01	0.000	.2898604	.5723865
autonregion	.1950642	.174969	1.11	0.266	-.1493957	.5395241
bordering	-.0358025	.0489331	-0.73	0.465	-.1321366	.0605316
pop	3.10e-07	2.90e-07	1.07	0.285	-2.60e-07	8.81e-07
_cons	.5179254	.334643	1.55	0.123	-.1408835	1.176734

## **List of Variables in the Group Dataset**

### **terrint** (dependent variable for model 1)

This variable is a measure of the occurrence and intensity of separatist conflict, on a scale from 0 to 3, measured for the period 1990-2003. Source: PRIO

### **separatism** (dependent variable in model 2)

This variable measures the strength of separatist sentiment among subnational groups, on a scale from 0 to 3, for the 1990s. Source: MAR.

### **africa**

Dummy variable, takes value 1 if group is African

### **lac**

Dummy variable, takes value 1 if group is Latin American.

### **gdp90**

Per capita GDP of the group's country, measured as of 1990 (same methodology as country dataset)

### **maleseced90**

Gross male secondary school enrollment, 1990.

### **oilandcoal**

Dummy variable. Takes the value of 1 if group's region harbors oil or coal. Note that this variable is coded using MAR description and author's knowledge for identification of the region where each group is located. Data source used for resources is USGS minerals map (last modified November 2002) located at:

<http://minerals.usgs.gov/minerals/pubs/country/maps/>

Also used University of Texas map website to access detailed maps of each country in order to determine which region any given group was located in:

<http://www.lib.utexas.edu/maps/>.

Groups are coded as having a resource when USGS defines a resource as undeveloped but significant (these resources appear in parentheses on the USGS maps). Groups that have these resources in their region are: Southerners in Chad, Somalis in Ethiopia, Nuba and Southerners in Sudan, Tonga in Zambia, Issaq in Somalia, and Diolas in Senegal.

The USGS site did not have a map for China. Authors substituted a 1983 CIA map of China's fuel, power, minerals and metals. This map is found on the UT map website at [http://www.lib.utexas.edu/maps/middle\\_east\\_and\\_asia/china\\_fuels\\_83.jpg](http://www.lib.utexas.edu/maps/middle_east_and_asia/china_fuels_83.jpg)

**ecdifabs**

Absolute value of group's economic differential score. Source: MAR.

**regional**

Dummy variable; 1 if the group has a region where it is predominant. Source: MAR.

**groupcon**

Group concentration, from 0 (widely dispersed) to 3. Source: MAR.

**belief**

Group's belief specificity compared to rest of the country. Source: MAR.

**custom**

Group's customs specificity compared to rest of the country. Source: MAR.

**lang**

Group's linguistic specificity compared to rest of the country. Source: MAR.

**race**

Group's racial specificity compared to rest of the country. Source: MAR.

**poppercent**

National population share of group. Source: MAR.

**language**

National linguistic diversity. Source: Alesina et al. 2002.

**diaspora**

Dummy version of MAR emigration variable, where all values greater than zero equal 1. Missing values imputed by authors to the best of our knowledge based on presence of kin abroad. Measured in previous period.

**dispersion**

Transnational dispersion of kindred groups. Source: MAR. Note: Countries separated by water distances of less than 200 miles should be considered "adjoining." Only the countries that are adjoining/neighbors of the country in question are coded.

**discrimination90**

Mean of economic and political discrimination variables, 1980-1990 (Source: MAR).

**allfailure90**

Same as in Country Dataset

**politrigh90**

Freedom House score on political rights for 1990 or first avail.

**autlost**

Index of Political Autonomy Grievances. Source: MAR.

Missing Values: -99

A composite index for groups who have lost autonomy or undergone a transfer of control from one country to another. Values range from "0" (no historical autonomy) to "6.0".

The index is constructed by adding the weights for "Magnitude of Change" and "Group Status Prior to Change," subtracting one, and dividing by the "Year-of Loss" weight. Weights were assigned using information coded below plus supplemental information (see Gurr, "Why Minorities Rebel", pg. 199, for more information).

These are the weights:

<i>Year of Loss of Autonomy or Change</i>		<i>Magnitude of Change</i>		<i>Group Status Prior to Transfer of Control</i>	
<i>Period</i>	<i>Weight</i>	<i>Type</i>	<i>Weight</i>	<i>Status</i>	<i>Weight</i>
1980-95	5	Loss of long-term autonomy	3	State or republic	4
1960-79	4	Loss of short-term autonomy (< 10 years) under colonial rule	2	Autonomous region or province, or autonomous people	3
1940-59	3	Transfer only centralized authority, religious or secular	1	Traditional	3
1900-39	2	Province in another state or colonial territory	2		
pre-1900	1	Part of larger segment of group	2		
		Autonomous but acephalous or	1		

**autonregion**

Auton2 variable from MAR. Scores 1 if group members govern one or more regions with at least limited autonomy, otherwise 0. Those that were missing were inferred by the authors—that includes codings for USSR and Czechoslovakia (they had been discontinued when the variable was created). Source: MAR.

**bordering**

Number of Segments in Adjoining Countries (MAR: numsegx). This variable contains the number of adjoining countries where there are segments of the group.

**pop**

Country population 1990s in 1,000s

**size**

Country size in square miles

**inpieces**

Dummy for country with non-contiguous territory