

## The Long-Run Effects of the Scramble for Africa

*American Economic Review*, forthcoming

Stelios Michalopoulos and Elias Papaioannou

---

### Replication Datasets (STATA .dta files).

- **Main Conflict (ACLED-UCDP) Dataset. (1997-2013)** "aer\_all2013.dta"
  - This is the main dataset to replicate all main paper tables linking ethnic partitioning and civil conflict as well as most appendix tables.
  - Tables 2-6; and Appendix Tables 2, 7, 9 and Appendix Tables 15-32.
- **Alternative Conflict (ACLED-UCDP) Dataset. (1997-2010).** "aer\_all2010.dta"
  - This dataset is otherwise identical to the main conflict dataset, but the ACLED data cover the period 1997-2010. [This is to account for the significant increase in reporting events in ACLED over 2010-2013.]
  - Appendix Tables 11, 12, 13, 14 and Appendix Tables 29 and 30
- **Main Ethnic Wars and Political Discrimination (EPR) Dataset.** "main\_epr.dta"
  - Table 7; and Appendix Tables 33, 34A,B (columns 1 and 2), 35 and 36.
- **Panel of Ethnic Wars and Political Discrimination (EPR) Dataset**  
"epr\_time\_series\_upload.dta"
  - Appendix Tables 34A, 34B; Columns 3-6
- **Main Individual-level Public Goods (DHS) Dataset.** "DHS\_replication\_Sept26th"
  - Tables 8-9; and Appendix Tables 37-40.
- **Ethnic Homeland Characteristics Dataset** "homelands825"
  - Table 1; and Appendix Tables 8 and 10.

We also include the concordance created to match the groups from the EPR to the Murdock Map. This correspondence can be found at "EPRtoMurdockMap\_Concordance.dta"

### Notes

- The folder includes the programs (Stata .do files) and data files (Stata .dta files) that replicate all tables in the main paper and all tables reported in the on-line Supplementary Appendix.
  - Please note that inside the ".do" files we include some useful comments.

- At some point perhaps one needs to “clear the estimates” (by typing “`estimates clear`”), as STATA may run out of memory.
- You need to install the STATA “`estout`” .ado so as to easily export tables.
- The folder includes a master execution “.do” program that runs all the programs with ACLED-UCDP (main conflict data). This execution file is named “`allacled_new`”. One can simply run this program as it nests many sub-programs (that are however also included in the folder).
- The .dta files include brief definitions for all variables. (Please type: “`describe all`”). The Data Appendix gives detailed variable definitions and data sources.
- In the beginning (and before running any regressions) please run the “`cgmreg.ado`” routine that allows getting double-clustered standard errors using the multi-way clustering method of Cameron, Gelbach and Miller. [A. Colin Cameron, Jonah B. Gelbach, and Douglass Miller. (2011). Robust inference with Multiway Clustering,” *Journal of Business and Economic Statistics*. April 2011, Vol. 29. No. 2. pp 238-249].
- As there is no routine [.ado file] to generate double-clustered standard errors with count data models (Negative Binomial and Poisson ML), for such models one has to run the program three times, so as to get the exact standard errors (which however are very similar to clustering simply at the ethnolinguistic family). First, please cluster at the country-level (`cluster(wbcode)`); Second, cluster at the group’s ethnolinguistic cluster (`cluster(cluster)`); Third, cluster at the intersection of a group’s country and ethnolinguistic cluster (`cluster(inter)`). Then one has to apply the following formula:  $SE_{double} = \text{SQRT}[SE_a^2 + SE_b^2 - SE_{a-b}^2]$ , where  $a$  and  $b$  denote the two dimensions, country and ethnolinguistic family in our setting, and  $a-b$  denotes the intersection. SQRT denotes the square root. See Cameron, Gelbach, and Miller (*JBES* 2011) and Cameron and Miller (A Practitioner’s Guide to Cluster-Robust Inference, *Journal of Human Resources*, 2014; in particular the formulas and method is Section V.A).
- In a few instances, the Negative Binomial ML model will not converge when we restrict estimation in the sample of country-ethnic homelands close to the national border (we use the median of distance to the border as a cut-off). Therefore in those specifications we exclude from the estimation country-ethnic homelands in countries with no variation on ethnic partitioning. These countries are Burundi, Djibouti, Swaziland, Madagascar, and Western Sahara. [The identifier variable for observations in these countries in the .dta file is “no”].
- Please set STATA “`matsize`” high (say 10,000), as the codes generate many estimations.
- For some tables, you need to install some specific STATA packages. Namely
  - For the test of medians in Appendix Table 10, please download and install “`qreg2`” routine that allows obtaining clustered standard errors for quantile regression. [if you are

not interested in the standard errors one can simply use “`qreg`” or directly compare the medians using the “`summarize`” command].

- For the spatial econometric specifications in Appendix Table 24, please download and install the “`spwmatrix`” routine that allows estimation of various spatial models.

## Structure of Execution (.do) Programs

---

---

### Ethnic Partitioning and Border Artificiality

[use "homelands825.dta"]

- **Main Paper. Section 3. Ethnic Partitioning and Border Artificiality**
    - Table 1.
  - **Supplementary Appendix. Section 2.**
    - Appendix Table 1 and Appendix Table 8
  - ***T1. Border (Ethnic Partitioning) Artificiality.*** [Run .do file T1.]
  - ***AT1. Appendix Table 1.*** Summary Statistics at the Ethnic Homeland Level [Run .do file T1.]
  - ***AT8. Appendix Table 8.*** Pre-colonial Ethnic Features (using data from Murdock (1967)) and Ethnic Partitioning [Run .do file T1.]
  - ***AT10. Appendix Table 10.*** Test of Means and Medians for the Main Conflict Measures at the Ethnic Homeland Level [Run .do file TA10.]
- 

### Ethnic Partitioning and Civil Conflict (ACLED & UCDP)

- **Main Paper. Section 4. Ethnic Partitioning and Civil Conflict**
  - Tables 2-4.
- **Main Paper. Section 5. Further Evidence**
  - Tables 5-6.
- **Supplementary Appendix. Section 3. Further Evidence and Sensitivity Analysis**
  - Appendix Table 2 (Summary Statistics), Appendix Table 7 (Correlation Structure), and Appendix Tables 11-30 (Robustness Checks),
  - Discussion in Main Paper, Section 4.6.
- **Supplementary Appendix. Section 4. Heterogeneous Effects**
  - Appendix Tables 31-32 (Heterogeneous Effects)
  - Discussion in Main Paper, Section 4.7.
- The .do file "results\_acled\_aer" runs all programs that estimate all specifications (in both the main part of the paper and the online Supplementary Appendix) linking ethnic

partitioning with civil conflict using ACLED and UCDP over 1997-2013 (Tables 2-6, Appendix Tables 2, 7, and 15-28).

- The programs [.do files] below are “nested” in the `results_acled_aer.do`.
  - Yet it is preferable to be run separately.
  - Perhaps at some point (ideally after running each execution file) you need to clear estimates as STATA may run out of memory [type “`estimates clear`”]
- 
- **T2. Ethnic Partitioning and Civil Conflict. Baseline Country Fixed-Effects Estimates.**  
**ACLED**
    - Panel A. Negative Binomial ML Estimates. [Run `.do file T2-A.`]
    - Panel B. Linear Probability Model (LPM) Estimates [Run `.do file T2-B.`]
  - **T3. Ethnic Partitioning and Civil Conflict Intensity. Country Fixed-Effects Estimates.**  
**ACLED** Duration, fatalities, deadly events [Run `.do file T3.`]
    - The program also estimates results in Appendix Table 17. [Poisson ML excl. outliers]
  - **T4. Ethnic Partitioning and Civil Conflict Aspects. Country Fixed-Effects Estimates.**  
**ACLED**
    - Panel A. Negative Binomial ML Estimates. [Run `.do file T4-A.`]
    - Panel B. Linear Probability Model (LPM) Estimates [Run `.do file T4-B.`]
  - **T5. Ethnic Partitioning and Conflict Actors. ACLED** [Run `.do file T5.`]
    - Panel A. Negative Binomial ML Estimates.
    - Panel B. Linear Probability Model (LPM) Estimates
  - **T6. Ethnic Partitioning and Civil Conflict Types. UCDP GED** [Run `.do file T6.`]
    - Panel A: State (Government Forces) Civil Conflict
    - Panel B: One-Sided Violence against Civilian Population
    - Panel C: Non-State Civil Conflict
      - Note in old STATA versions there may be some convergence issues; please run the code dropping outliers or countries with no variability in non-state conflict.
  - **AT2. Summary Statistics. ACLED.** [Run `.do file TA2.`]
  - **AT7. Correlation Structure - Main Conflict Variables (ACLED and UCDP).** [Run `.do file TA7.`]
  - **AT9. "Balancedness Tests." Ethnic Partitioning and Geographic Characteristics within Countries** [Run `.do file TA9.`]

- **AT10. Preliminaries. Ethnic Partitioning and Conflict at the Ethnic Homeland Level**  
[Run .do file TA10.]
  - Test of Means and Medians for Main Civil Conflict Measures (ACLED & UCDP)
  - This program has to be run at the "homelands825" database.
  - Please install STATA "qreg2" routine that allows getting clustered standard errors in quantile regression.
  - In cases where the median is 0 for both split and non-split ethnicities, then qreg2 doesn't run in some computers. You can simply use the "qreg" command in these cases.
- **AT11. Ethnic Partitioning and Civil Conflict. Baseline Country Fixed-Effects Estimates over 1997-2010. ACLED.**
  - Identical to Table 2, but estimated using ACLED data over 1997-2010.
  - Just run execution files "T2-A.do" and "T2-B.do" using the dataset (STATA .dta file) "aer\_all2010.dta."
- **AT12. Ethnic Partitioning and Civil Conflict Intensity. Baseline Country-Fixed-Effects Estimates over 1997-2010. ACLED.**
  - Identical to Table 3, but estimated using ACLED data over 1997-2010.
  - Run execution file "T3.do" using the dataset (STATA .dta file) "aer\_all2010.dta."
- **AT13. Ethnic Partitioning and Main Aspects of Civil Conflict over 1997-2010. ACLED.**
  - Identical to Table 4, but estimated using ACLED data over 1997-2010.
  - Just run execution files "T4-A.do" and "T4-B.do" using the dataset (STATA .dta file) "aer\_all2010.dta."
- **AT14. Ethnic Partitioning and Conflict Actors over 1997-2010. ACLED.**
  - Identical to Table 5, but estimated using ACLED data over 1997-2010.
  - Run execution file "T5.do" using the dataset (STATA .dta file) "aer\_all2010.dta."
- **AT15. Ethnic Partitioning and Civil Conflict. Alternative Estimation Techniques**
  - Panel A. Conditional Negative Binomial ML Estimates. [Run .do file TA15-A.]
  - Panel B. Fixed-Effects Poisson ML Estimates, excl. Outliers. [Run .do file TA15-B.]
  - Note. For this table and all subsequent tables we revert to the baseline sample that covers the period 1997-2013.
- **AT16. Ethnic Partitioning and Main Aspects of Civil Conflict. Alternative Estimation Techniques**

- Panel A. Conditional Negative Binomial ML Estimates. [Incl. in the program TA-15A.  
Run `.do file TA15-A.`]
- Panel B. Fixed-Effects Poisson ML Estimates, excl. Outliers. [Incl. in the program  
TA15-B. Run `.do file T15-B.`]
- **AT17. Ethnic Partitioning and Civil Conflict Intensity. Poisson Country Fixed-Effects Estimates, excl. Outliers.** Duration, fatalities, deadly events [Run `.do file TA17.`]
- **AT18. Ethnic Partitioning and Civil Conflict Intensity. Not Accounting for Spillovers.** [Run `.do file TA18.`]
- **AT19. Ethnic Partitioning and Civil Conflict Intensity. Alternative Measure of Ethnic Partitioning.** [Run `.do file TA19.`]
- **AT20. Ethnic Partitioning and Civil Conflict Intensity. Controlling for Unobservables. Distance to the Border. 3rd-order Polynomial.** [Run `.do file TA20.`]
- **AT21. Ethnic Partitioning and Civil Conflict Intensity. Controlling for Unobservables. Distance to the Border. 3rd-order Polynomial.** [Run `.do file TA21.`]
- **AT22. Ethnic Partitioning and Civil Conflict Intensity. Ethnic Family Fixed Effects Specifications.** [Run `.do file TA22.`]
- **AT23. Ethnic Partitioning and Civil Conflict Intensity. Dropping Iteratively a Different African Region.** [Run `.do file TA23.`]
- **AT24. Ethnic Partitioning and Civil Conflict Intensity. Accounting for Spillovers with Spatial Models.** [Run `.do file TA24.`]
  - Need to download and install the “spmlreg”, “spwmatrix”, and “splayvar” STATA modules to run the spatial models.
  - Please run this program in the end (after running all other .do files).
  - This program is somewhat slow (especially for the generalized spatial model).
- **AT25. Ethnic Partitioning and Civil Conflict Intensity. Accounting for Spillovers at the Country Level and at the Ethnic Family Level** [Run `.do file TA25.`].
- **AT26. Ethnic Partitioning and Civil Conflict Intensity. Accounting for Pre-colonial Conflict and Political Centralization.** [Run `.do file TA26.`]
- **AT27. Ethnic Partitioning and Civil Conflict Intensity. Accounting for Regional Development** [Run `.do file TA27.`]
- **AT28. Ethnic Partitioning and Conflict Actors. Accounting for Regional Development** [Run `.do file TA28.`]
- **AT29. Ethnic Partitioning and Civil Conflict. Accounting for Measurement Error in the Civil Conflict Databases** [Run `.do file TA29.`]

- This execution file has to be run on the dataset acled2010, as UCDP GED 1.5 stops in 2010. [The results however are similar even if the program is run in the main data-set, acled2013].
- **AT30. Ethnic Partitioning and Civil Conflict. Accounting for Measurement Error in the Civil Conflict Databases** [Run `.do file TA30.`]
  - This execution file has to be run on the dataset acled2010, as UCDP GED 1.5 stops in 2010. [The results however are similar even if the program is run in the main data-set, acled2013]

---

### Heterogeneous Effects of Ethnic Partitioning on Civil Conflict (ACLED)

- **Supplementary Appendix Section 5.**
    - Discussion in Main Paper, Section 4.7.
  - **AT31. Heterogeneous Effects by Country-Ethnicity Features** [Run `.do file TA31.`]
  - **AT32. Heterogeneous Effects by Country Features** [Run `.do file TA32.`]
- 

### Ethnic Partitioning and Ethnic Power Relations

- For all specifications please use `"epr_main.dta"`.
- To generate all tables in the main part of the paper and the cross-sectional ones in the Supplementary Appendix) with the EPR data please run STATA execution file `"epr_all.do"`.
- Please make sure that you have installed or run before estimating the EPR Tables the `"cgmreg.ado"` that allows getting double-clustered standard errors with the multi-way clustering method of Cameron, Gelbach and Miller. [A. Colin Cameron, Jonah B. Gelbach, and Douglass Miller. (2011). Robust inference with Multiway Clustering," *Journal of Business and Economic Statistics*. April 2011, Vol. 29. No. 2. pp 238-249]
- **Main Paper, Section 6**
  - **Table 7. Ethnic Partitioning and Political Violence**
    - Ethnic Civil Wars
    - Political Discrimination from National Government



- Ordered Political Violence
  - **Supplementary Appendix, Section 5.**
    - **Appendix Table 33.** EPR Descriptive Patterns
    - **Appendix Table 34.** Ethnic Partitioning, Exclusion from Central Government and Ethnic War Incidence and Onset. For the panel estimates in Columns 3-6 of both Panels please use the “epr\_times\_series\_upload.dta” and the associated do file: AT34\_panel.
    - **Appendix Table 35.** Sensitivity Analysis. Alternative Models Linking Ethnic Partitioning to Political Violence.
    - **Appendix Table 36.** Sensitivity Analysis. Alternative Index of Ethnic Partitioning.
    - **Appendix Table 37.** Sensitivity Analysis. Excluding Each Time a Different African Region
    - **Appendix Table 38.** Sensitivity Analysis. Excluding Each Time a Different African Region and Employing an Alternative Index of Ethnic Partitioning.
- 

## Ethnic Partitioning and Individual Well-Being

- For all specifications please use “dhs\_main.dta”.
- To generate all tables (in the main part of the paper and the Supplementary Appendix) with the EPR data please run STATA execution file “dhs\_all.do”.
- Please make sure that you have installed or run before estimating the DHS Tables the “cgmreg.ado” that allows getting double-clustered standard errors with the multi-way clustering method of Cameron, Gelbach and Miller. [A. Colin Cameron, Jonah B. Gelbach, and Douglass Miller. (2011). Robust inference with Multiway Clustering,” *Journal of Business and Economic Statistics*. April 2011, Vol. 29. No. 2. pp 238-249]
- **Main Paper, Section 7**
  - **Table 8.** The Long-Run Effects of Ethnic Partitioning on Individual Well-Being and Education. DHS Data. Baseline Estimates)
  - **Table 9.** The Long-Run Effects of Ethnic Partitioning on Individual Well-Being and Education. DHS Data; Location and Identity.
- **Supplementary Appendix, Section 7.**
  - **Appendix Table 39.** Descriptive Patterns and Summary Statistics

- **Appendix Table 40.** The Long-Run Effects of Ethnic Partitioning on Individual Well-Being and Education. DHS Data. Enumeration-area (Village) Fixed Effects Estimates
- **Appendix Table 41.** The Long-Run Effects of Ethnic Partitioning on Individual Well-Being and Education. DHS Data. Location and Identity. Looking at "Movers" & "Non-Movers"
- **Appendix Table 42.** The Long-Run Effects of Ethnic Partitioning on Individual Well-Being and Education. DHS Data. Examining Persistence.