

Notes on analyzing ethnic groups and ethnic conflict

James D. Fearon
Stanford University

October 3, 2019

Authorization for this talk

“... participants will be allowed to present not only recent and unpublished works but also old works that they want to somehow revisit or put in the perspective of newly accumulated evidence and current realities and interpretative schemes.”

–Ruben, Jean-Philippe, and Shlomo

Summary

Agenda, argument

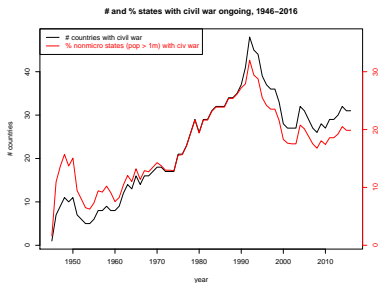
1. Why David and I moved in 1990s from ethnic groups to countries as primary units of empirical analysis.
 - ▶ The difficult problem of ethnic groups being endogenous to same causal processes that produce “stability and development.”
2. It doesn't make much sense to try to estimate “average treatment effects” of historically determined things like fractionalization, polarization, or degrees of cultural difference.
3. Makes more sense to study the **mechanisms** (causal logics) that lead to the politicization of ethnicity in some circumstances. And the mechanisms that relate politicized ethnicity to important political and economic outcomes.
4. Deutsch/Gellner/Anderson's modernist theory(ies) of nationalism as an example/model.
5. Some very quick empirics showing failure of cultural distance measures to predict ethnic group having an armed group fighting separatist civil war.

Background: End-of-Cold-War effects

- ▶ When David and I started collaborating in the mid 1990s, our interest and focus was on **ethnic conflict**.
 - ▶ Violent break up of Yugoslavia. Some post-Soviet conflicts in E. Europe and FSU.
 - ▶ We had also common interests in African politics and conflicts (Somalia, Liberia, Sierra Leone, Rwanda . . .).
- ▶ Recall that 1990s saw lots of efforts to frame/conceptualize the new era of post-Cold War int'l politics.
 - ▶ Major contenders early on included “resurgence of suppressed ancient ethnic hatreds” (Kaplan, others) and “clash of civilizations” (Huntington).

Background

- ▶ Seemed to us (and others) an important area that lacked for basic empirical and theoretical work.
- ▶ Civil/ethnic conflict had been pretty systematically ignored in US social science during Cold War.



Background: First tried to study ethnic groups and ethnic conflict

- ▶ We started out focused on and thinking about **ethnic groups**.
 1. “Explaining interethnic cooperation” (*APSR* 1996). Social matching game model of inter- and intra-group interactions.
 2. “Weak states, rough terrain, and large-scale ethnic violence since 1945” (long, unpublished conference paper, 1999 APISA).
- ▶ Empirical part of (2) used Ted Gurr’s Minorities at Risk data to ask what characteristics distinguished ethnic groups that had been involved in significant violence from those that had not.

Background: Gurr's Minorities at Risk (MAR) data

- ▶ Gurr conceived of MAR as a project to enable forecasting of events of genocide, repression, cultural elimination of minority groups.
- ▶ Gurr et al coded 100s (!) of variables for 268-350 “minorities at risk” in around 110 countries.
- ▶ “At risk” said to mean that group either
 1. suffers “discrimination” relative to other groups in country, or
 2. is “disadvantaged from past discrimination,” or
 3. is an “advantaged minority being challenged,” or
 4. is “mobilized,” meaning that “the group (in whole or part) supports one or more political organizations that advocates greater group rights, privileges, or autonomy.”

Background: Gurr's Minorities at Risk data

- ▶ MAR coded two broad forms of ethnic violence: Rebellion (with the state) and Communal Conflict (between groups).
- ▶ Surprising observation in our 1999 paper: *Very little CC relative to Rebellion, and most major CC is in context of Rebellion.*
- ▶ → Fighting between state and armed groups “representing” ethnic group(s) is **by far** most significant form of “ethnic violence” in terms of death and destruction.

Background: Gurr's Minorities at Risk data

- ▶ MAR groups and rebellion:
 - ▶ 55% of the 268 were coded as having been involved in some degree of armed rebellion against state between 1980-1995.
 - ▶ 35% at level of “small-scale guerrilla war” or higher!
- ⇒ There is major selection on the dep var in MAR list!
- ▶ Groups included on the basis of “at risk,” which is judged implicitly/explicitly by conflict experience of group or expectation of factors correlated with conflict propensity.
- ▶ More on this selection issue in a minute.

Background: Our 1999 paper's empirical findings

- ▶ The empirical half of our 1999 paper used the 268 “Phase III” MAR groups, some MAR variables, and some new variables that we coded.
- ▶ Some findings foreshadowed some of the main results in FL 2003, “Ethnicity, insurgency, and civil war” – a country-year analysis of civil war onset, ethnic or not.
- ▶ But 1999 paper also had some empirical results at level of ethnic group characteristics.

Background: Our 1999 paper's empirical findings

Summary:

- ▶ Ethnic group more likely to be in rebellion with state after 1980 if group
 1. Lives in a low income country, with lower growth rate before 1980.
 2. Has a regional (geographically concentrated) base, lives in rough terrain, is not urban, has co-ethnics that dominate a neighboring state.
- ▶ **Not** related to prob of significant violence with state:
 1. Linguistic or religious distance from dominant group in state!
 2. Level of democracy of country (controlling for income)!
 3. MAR's measures of cultural and economic discrimination against the group!

Background: But then we gave up on ethnic group level analysis!

- ▶ However, by the time presented 1999 paper at APSA, we were already giving up on ethnic-group-level analysis.
- ▶ Draft of “Ethnicity, insurgency, and civil war” (*APSR* 2003) presented for the first time in April 2000, at Laboratory in Comparative Ethnic Processes (LiCEP) meeting.
 - ▶ Dep var = 1 if country-year saw onset of any civil war.
 - ▶ Country-year panel, 1945-1999.
- ▶ Why abandon ethnic group/ethnic violence focus?

“Ethnic violence” is a problematic category. And ethnic groups are “endogenous.”

1. We gradually realized that civil war was a conceptually clearer, easier category to deal with than ethnic violence.
 - ▶ Also, the data were telling us that the interesting questions were about civil war more broadly – organized non-state armed groups fighting with or over the state.
2. We had spent **much** time puzzling over how to conceive of the population of ethnic groups from which MAR groups were some kind of sample.
 - ▶ Decided that the conceptual/theoretical issues here were **much** harder, much trickier than saying, Our unit of analysis is given by the formalized set of UN system country boundaries.

Selection bias in MAR

- ▶ Why not just go with MAR groups? (and perhaps we should have tried to publish that damn paper)
- ▶ We were very concerned about **selection bias** due to the “at risk” criteria. Eg:
 - ▶ What if variation in our measures of cultural difference between minority and dominant group was heavily truncated by restriction to conflict-prone groups?
 - ▶ eg, what if more peace-prone minorities, omitted from MAR, are much more similar to dominant group on average?

Selection bias in MAR

- ▶ *But we couldn't figure out how to conceive of and code "the population" of ethnic groups.*
- ▶ This is necessary to do *any* empirical analysis at the level of ethnic groups (or ethnic group-years).
- ▶ The problem is that groups – and not just ethnic groups – are social constructs (beliefs and related actions) that change over time in a dynamic interplay with political, social, and economic events/developments.

Ethnic groups are “endogenous.”

We were acutely aware of (eg)

- ▶ The Pol Sci literature on “situational ethnicity.” Filled with wonderful eg’s of the same individual(s) conceiving of their ethnic group membership differently in different social/political contexts.
- ▶ The “modernist” literature on origins of nationalism, which is based on the observation that people around the world did not systematically conceive of cultural groups as the proper basis for political mobilization and political boundaries until the last ~250 years. Much more recently in most of the world.
- ▶ Fact that most current ethnic groups in Africa simply did not exist as such in the minds of their (later) members in the pre-colonial period. Most emerged in their present form as a result of deliberate political and social mobilization, colonial and post-colonial state policies in last 150 years.

Ethnic groups are “endogenous.”

- ▶ Variability and recent-ness of modern ethnic/national groups would not be an issue if this variation was orthogonal to our political and economic outcomes of interest.
- ▶ But not so! Violent conflict, economic development, and “strong state institutions” undoubtedly can causally affect both peoples’ conceptions of ethnic groups, and also how we as analysts code these.

Ethnic groups are “endogenous.”

- ▶ eg: relatively strong French and Prussian states deliberately created sense of ethnic/national homogeneity in France and Germany in 19th C.
- ▶ eg: economic success, and Somalia vs Botswana.
 - ▶ Somalia was seen – plausibly – as the most ethnically homogeneous country in subSaharan Africa until it fell apart due to fighting organized along clan lines.
 - ▶ Botswana also seen and coded as quite ethnically homogeneous, even though Tswana have multiple subtribes, analogous to Somalia clans.
 - ▶ Plausible that econ/political success of Botswana causes academic and local perception of ethnic homogeneity of Tswana, whereas econ/political failure has caused local and academic perception of “ethnic” divisions in Somalia.

Ethnic groups are “endogenous.” The Ethnic Power Relations dataset.

- ▶ EPR, a more ambitious and academically successful successor to MAR, shows the same problems coming from fact that “ethnic groups” are endogenous to the outcome variables that we want group characteristics to explain.
- ▶ Lars-Erik Cederman, Andreas Wimmer, and collaborators organized expert coding of 892 “politically relevant” ethnic groups in 181 countries, 1946-2017.
 - ▶ “Politically relevant” if “at least one signif political actor claims to represent the group in national political arena, or if members are systematically and intentionally discriminated against in the domain of politics” (*ASR* 2009).

Ethnic groups are “endogenous.” The Ethnic Power Relations dataset.

Some weird and problematic implications of the “politically relevant” groups approach:

1. Some highly ethnically diverse countries get coded as highly homogeneous: Eg,

Country	EPR group	Population share
Papua New Guinea	Papua New Guineans	.966
	Bougainvilleans	.034
Tanzania	Mainland Africans	.964
	Shirazi (Zanzibar Africans)	.005
	Maasai	.013

Ethnic groups are “endogenous.” The Ethnic Power Relations dataset.

2. Conflict and political instability **clearly** causes definition of, and sometimes over-time changes to, EPR group list.
 - ▶ eg of PNG. Bougainville is an island that is itself highly ethnically diverse (e.g., at 26 distinct languages from three completely different language families). The reason for the EPR coding is the political and armed conflict between Bougainville and the PNG capital (1989-98).”
 - ▶ Likewise case of Zanzibar for Tanzania.

Ethnic groups are “endogenous.” The Ethnic Power Relations dataset.

- ▶ Another interesting example, Liberia.

For 1946-1980		For 1981-2007	
Americo-Liberians	.02	Americo-Liberians	.02
Indigenous Peoples	.98	Gio	0.08
		Mano	0.07
		Krahn (Guere)	0.05
		Mandingo	0.017

- ▶ The reason for the change is that in April 1980, Master Sergeant Samuel Doe (Krahn) carried out a violent coup against William Tolbert (Americo-Lib). EPR considers that this caused Gio, Mano, Krahn, and Mandingo (but not many other mainland ethnic groups!) to become “politically relevant.”

Ethnic groups are “endogenous.” The Ethnic Power Relations dataset.

- ▶ How do you even think about a panel data set in which the *units of analysis* in one year depend on values of the dep var in previous years, or on expectations of dep var values in future years?
- ▶ I find this very confusing.

Ethnic groups are “endogenous.” The Ethnic Power Relations dataset.

Problematic implications of the “political relevant groups” approach, cont.:

3. Finally, EPR group list is still heavily selected on conflict or conflict propensity:
 - ▶ 27% of 892 EPR groups have at least one year of civil conflict using Uppsala ACD codings.
 - ▶ Compare to 45% of MAR groups, and 14% of groups in my 2003 JoEG list (more in a minute).
 - ▶ Moreover, EPR heavily overrepresents very small (< 1% country population) groups that were involved in armed conflict at some time.

Ethnic groups are “endogenous.” What to do?

- ▶ In making these criticisms of MAR and EPR, I’ve been implicitly assuming that these lists do not properly represent “the population” of ethnic groups across countries.
- ▶ But how to do this?

Ethnic groups are “endogenous.” What to do?

- ▶ iirc, David sometimes argued that in principle one would want to list **all hypothetically possible** “ethnic groups” in a country.
 - ▶ eg, list not only Hispanics in the US, but ALSO Mexican Americans, Cuban Americans, El Salvadorans, etc. Likewise Asians.
 - ▶ More extreme: Allow hypothetical groups including people currently coded as White and Hispanic, etc.
- ▶ Point is that if ethnic groups are endogenous products of history, politics, social interaction, then the ones we see are a tiny subset of all that could have or could still occur.

Ethnic groups are “endogenous.” What to do?

- ▶ My recollection is that in comparison to David, I was more optimistic that it was both possible and desirable to construct a meaningful, useful list of “the population of ethnic groups” across countries.
- ▶ I thought one key was to avoid (at least in principle) the idea of “political relevance,” as in MAR, EPR, and an influential article by Dan Posner (*AJPS* 2004).
- ▶ Instead, attempt to code **socially relevant** ethnic categories – how people imagine the ethnic landscape in everyday life. Thus covering whole country, whether or not group is “politically relevant” (endogenous question of interest).

Ethnic groups are “endogenous.”

- ▶ In “Ethnic mobilization and ethnic violence” (Oxford Handbook chapter, 2006) I distinguished between ethnicity being
 1. **socially relevant** when people notice and condition their actions on ethnic distinctions in everyday life, and
 2. **politicized** (or politically relevant) when political coalitions are organized along ethnic lines, or when access to political or economic benefits depends on ethnicity.
- ▶ Ethnicity can be socially relevant in a country without it being much politicized, and the degree to which ethnicity is politicized can vary across countries and over time.
- ▶ Politicization of ethnicity seems like interesting and important variation worth documenting and studying systematically.

Ethnic groups are “endogenous.”

- ▶ Of course politicization at time t can affect social relevance at time $t + 1$! So not saying that this is a perfect solution to issues raised by fact of endogenous ethnicity.
- ▶ Fwiw though my intuition is that the social understandings of ethnic groups don't change that fast on average, even if in many places they also have *matryoshka* doll structures.

Ethnic groups are “endogenous.” 2003 J of Econ Growth list.

- ▶ “Ethnic and cultural diversity across countries (*J. of Econ Growth* 2003) tried to code about 160 countries ethnic landscapes according to how people in country imagined it.
- ▶ If you ask someone, “What are the ethnic groups in this country?” what would the modal answer be?
- ▶ This is still a highly problematic and unclear for some countries. eg India.
- ▶ Also, very difficult, maybe impossible to identify, distinguish, and measure size of **all** very small ethnic groups. I restricted attention to groups with at least 1% of country population.
- ▶ List has ~ 858 groups in 166 countries. Leaves a lot to be desired and I would like in principle to return to it.

ATEs vs mechanisms

- ▶ Current standard thinking: Endogeneity problem? You need an instrument!
 - ▶ Ethnic groups are endogenous so identify some part of cross-nat'l variation in diversity or group characteristics that was determined a really long time ago.
 - ▶ Use this variation to estimate contemporary ATE of diversity (eg).
- ▶ I have problems with this approach in general, even if one grants the usual assumptions for a valid instrument.

ATEs vs mechanisms

- ▶ Lack time and ability to work through these concerns in this talk.
- ▶ But in a nutshell:
 - ▶ ATE is only well defined and meaningful if we can say what it would mean if treated units had not been treated, and control units had been treated.
 - ▶ eg, in drug trial, control indiv would have gotten drug, and treated indiv would have gotten placebo. Easy.

ATEs vs mechanisms

- ▶ Say we consider Tanzania “treated” by high ethnic diversity. How do we imagine counterfactual Tanzania with low ethnic diversity?
- ▶ There are lots of ways this could have happened, historically. But *different paths to low-diversity Tanzania could imply different treatment effects*.
 - ▶ eg, lots of subtribes of larger mega-tribe \neq all became Muslims which changed social structure \neq homogenized by state set up by successful foreign invader group.
- ▶ With historical instruments approach, we are getting some sort of hard-to-intepret average effect, implicitly averaging over different implicit assignment paths in our historical sample.
- ▶ And if/when we estimate a LATE in this approach, it remains mysterious as to what is giving rise to it, ie, what is the causal mechanism(s) (just as in case of drug trial).

An alternative approach: Study specific mechanisms (or paths) that politicize ethnicity

Modernist theories of Nism (Deutsch, Gellner, Anderson) all addressed the following central question:

- ▶ How did we go from a world where ethnic/cultural groups were not seen as basis for political legitimacy and boundaries to a world where this is the norm?



An alternative approach: Study specific mechanisms (or paths) that politicize ethnicity

- ▶ Their answer super briefly:
 - ▶ In agrarian world there was \approx zero social mobility possible for 99%.
 - ▶ If lord in castle spoke different language to his wife, this was completely irrelevant for the peasants.
 - ▶ With rise of modern state, trade, econ growth, social mobility prospects increase, but become conditioned on language/shared culture.

An alternative approach: Study specific mechanisms (or paths) that politicize ethnicity

- ▶ DGA hypothesis: Separatist Nism more likely to develop when upward mobility prospects of a set of people are blocked by a shared ascriptive trait(s) (esp language or culture).
- ▶ In other words, when you have **discrimination** by cultural characteristics in politics, administration, and/or labor markets.
- ▶ Deutsch and Gellner also argued that probability of discrimination and a separatist nationalist movement was increasing in **extent of pre-existing cultural differences**.
 - ▶ They all contrast the experience of France (Romance languages, easier to assimilate to Paris) to that of Austria-Hungary.

An alternative approach: Study specific mechanisms (or paths) that politicize ethnicity

- ▶ This hypothesis hasn't been broadly tested, so far as I know.
- ▶ Here is a super quick and dirty take, using my (updated) 2003 JoEG group list.
- ▶ I coded which groups had armed organizations fighting separatist civil wars, 1946-2017, from David's and my civil war list.
- ▶ I brought in our measure of linguistic distance, coding distance of separatist group from
 1. plurality group in country, and
 2. random individual in country.

Cultural distance and separatist conflict, group-level patterns

- ▶ 51 of 859 groups (5.9%) involved in a separatist FL civil war at some time in 1946-2017.
- ▶ This is about 7.3% of all non-plurality groups (which almost never are separatist).
- ▶ $30/166 = 20\%$ countries with a $> 1\%$ ethnic group fighting separatist war at some time.
- ▶ Worth noting that there is a surprising number of $< 1\%$ groups that fight separatist conflicts. So omitted here. If coded “all” $< 1\%$ groups, would be a very small share of those, of course.

Cultural distance and separatist conflict, group-level patterns

DV = 1 if group involved in separatist war

grp pop share	0.134 (0.129)	0.138 (0.113)	0.141 (0.131)	0.147 (0.116)
lang distance (0-1) from plurality grp	0.019 (0.031)	0.0003 (0.030)		
relig distance (0-1) from plurality group			0.036 (0.031)	-0.010 (0.027)
constant	0.049** (0.022)		0.036 (0.023)	
Observations	694	694	663	663
Country fixed effects?	N	Y	N	Y

Notes: OLS. se's clustered on country. Plural groups omitted. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Cultural distance and separatist conflict, group-level patterns

- ▶ Not shown (yet), but interaction of linguistic and religious difference does predict armed separatism of group in terms of within country comparisons (ie country fe's).
- ▶ Consistent with some work by Joel Selway on overlapping cleavages and conflict.