

Nation-building through the Recognition of Minorities*:

The Effect of Multiethnic Regime Engineering on Identity Transmission During and After Albania's Communist Dictatorship

Pre-Analysis Plan
Mary Shiraef

Political Science, University of Notre Dame, IN, USA.

*This is a working draft. Before sharing or citing, please email me for the latest version: mshiraef@nd.edu.

August 18, 2022

Contents

1	Overview	2
2	Conceptualization, Measurement, and Theory	4
2.1	Independent Variable: Multiethnic regime engineering (MRE)	4
2.2	Case Fit: Communist Albania as a case of MRE	6
2.3	Sampling Frame: Treatment and Control Group Descriptions	6
2.4	Dependent Variable: Ethnopolitical Identity Compliance	9
2.5	Pre-existing (Family Identity) Conditions	10
2.6	Theory	11
3	Specified Hypotheses	12
4	Relevant Background and Scope of Study	14
4.1	Literature	14
4.2	Hoxha's Experiment	15
5	Design	16
5.1	Model	18
5.2	Parallel Trends	19
5.3	Sample size and selection	19
6	Dataset Description Overview	19
6.1	1930-45 Civil Registry Records	19
6.2	Cemetery data Overview	20
6.3	Addressing Weaknesses	20
6.4	Linking the Data	22
7	Discussion	22
8	Conclusion	22
A	Albanian Civil Registry, 1930-45, Dataset Description	25
B	Villages Names Over Time	27
C	Albanian Cemetery Dataset Description	27

1 Overview

Were local-level national identity disputes quelled by communist regimes—only to re-emerge generations later? Or did communist-era policies toward minority groups more decisively engineer communist and post-communist identity landscapes? Contrary to the assumption that authoritarian regimes do not tolerate diversity, we understand from historical research, i.e., (Martin, 1998; Suny, 1993) that Soviet-era border-making departed from the “melting pot” model of the Western world by introducing identity recognition policies toward minority identities within regimes’ borders—alongside their national identity building programs. These multiethnic policies allowed for and even resourced ethnic groups with the long-term goal of incorporating people with diverse backgrounds into the Soviet socialist scheme. They took an ethnoterritorial approach to border-making within regime’s international borders—which meant that after communists established local control, people often found themselves living in a territory with a newly assigned identity from the one(s) their family carried before. (1) Some people’s newly recognized identities were correctly assigned, i.e., in correspondence with the trend of their family history. (2) Others were incorrectly assigned, and (3) others were newly assigned a homogeneous identity despite not having identity features in either the direction of the majority or a recognized minority identity. More than 300 million people were directly affected by these policies in the first half of the 20th century. Yet, little is known about the effects of these policies on communist or post-communist identity transmission processes.

Enver Hoxha, the leader of Albania’s communist period, 1945-90, imitated Stalin’s approach toward minority groups. In southern Albania, he recognized Greek and Macedonian minorities at the respective borders of Greece and then-Yugoslavia (now North Macedonia). The identity features of the population prior to communism have been widely disputed—especially in relation to the Greek minority. Recent availability of micro-level data on this population affords an opportunity to map the social, linguistic, religious, and occupational identities prior to communism, 1890-1945, and see the changes which occurred during communism, 1945-1990. The multiethnic approach the regime took offers this study a quasi-experimental design with two treatment groups: 1) those deemed to be in the Greek minority and 2) those who were excluded despite prior open access to transnational identity-building and were instead, assimilated into the Albanian ethnic majority group. The control group for the study is derived from the population across the Albania-Greece border in today’s northern Greece. They experienced the rise of communism, similar to the treatment groups, but after 1945, experienced only assimilation into the Greek national identity. The key outcome I examine is long-term compliance of one’s identity transmission process with the state’s ethnic and/or political identity designation (hereafter, ethno-political identity compliance) versus failure to comply via memory attachment to one’s family’s pre-communist past and/or transmission of anti-socialist attitudes.

In line with much of post-Soviet literature, I hypothesize that minority recognition policies worked to

foster ethnic identities but failed politically to unify ethnic groups toward socialist ideology. Beyond the literature, I specify the conditions under which the policy failed politically. Specifically, I hypothesize that the policy formed respectively the ethnic “Greek minority of Albania” as well as compliance with the Albanian majority ethnic group, initially in the early communist period, 1945-60—but failed on political and social dimensions in the long-term among those with family histories incorrectly assigned a new identity in the opposite direction of their ancestors.

To measure ethnic identity compliance in Albania’s early communist period, 1945-1960, I use the number of people whose first names match pools of names linked predominantly to either linguistic identity (Albanian/Greek) in the decade prior to the study’s pre-period, 1930-45. For ethnic but not political identity compliance, I use this first name measure paired with village-level electoral turnout proportions. Hoxha’s party counted electoral turnout as a direct signal of a village population’s support for socialism. For long-term ethnic identity transmission compliance, I use the number of people who named all their children with names matching the state-assigned ethnopolitical designation and were memorialized by their children with use of the associated linguistic script (Albanian/Greek) on their graves. For long-term political identity transmission compliance, I use the number of village-level votes for parties with socialist platforms after communism, 1990-.

To test my hypotheses, I apply a difference-in-differences (DD) approach which leverages the year 1945, the year of 1) Hoxha’s sudden and forceful communist takeover, 2) his application of a territorially designated Greek minority, and 3) his closure of the border with Greece. My data sources are Albania’s first national civil registry, which documented individual-level information for every household 1930-45; Albania’s communist-era civil registry, which continued to collect information with the same data structure, 1945-60; cemetery records, 1800-now, which depict the process of legacy-building at the family level; and voting records, 1945-now. I link these records, using first names, surnames, parents’ first names, village, and birth years—which allows me to compose repeated measurements of identity for the same individuals before, during, and after communism. To address the capacity for serial correlation resulting from the same individuals over time, I cluster the standard errors of my analysis at the village level or at the family level when possible—reporting the standard errors of both. I estimate the standard errors for my DD model through ordinary least squares.

Soviet-era identity engineering policies have been covered extensively from a historical perspective, but their long-term effects, if any, are under explored. The vantage points of pre-communist information as a baseline and a strict border closure that lasted for more than 40 years provide a unique opportunity to causally identify the long-term effects of communist minority identity engineering on political identities. This study’s results will provide empirical clarity to a topic long disputed by historians and policymakers in both national and international contexts regarding “the Greek minority of Albania”—as well as one that remains a sticking point within ongoing negotiations to transition Albania into the European

Union (EU). More broadly, this study contains insights for minority identity disputes in post-communist states in Eastern Europe, especially on the role of historical treatment of minorities in post-authoritarian democratization processes.

2 Conceptualization, Measurement, and Theory

2.1 Independent Variable: Multiethnic regime engineering (MRE)

Multiethnic regime engineering (MRE) is the state-led dual process whereby minority group rights are politically guaranteed for a specified minority population group within an ethnoterritorial border, usually through affirmative action, cultural rights (provisions for native language schools, holidays, newspapers, and theatres, etc.), and national education in a minority language—alongside a similarly resourced national identity program for a majority group. The term combines from existing literature (1) the ethno-political motivation behind the term “multiethnic regimes” (Ş. Aktürk, 2011) and (2) an explicitly top-down framework, i.e., a “diversity regime” which links identity management to power and is agnostic to the identity claims made by the impacted group(s) (Phillips & Reus-Smit, 2019). Multiethnic regimes are regimes or policies that permit and institutionalize membership for more than one ethnic group, as opposed to “monoethnic regimes,” which limit membership to one ethnicity, i.e., Nazi Germany, or “anti-ethnic regimes,” which permit membership to more than one ethnic group but deny expressions of distinct ethnicities in the state apparatus, i.e., Turkey, Greece, or pre-1960 United States. The ideal type “multiethnic regime” is the former Soviet Union, earning the title of an “affirmative action empire” (Martin, 2001) with its *korenizatsiya* (indigenization) policies. *Korenizatsiya* introduced the three main components of MRE to populations within the Soviet Union: (1) newly enforced ethnoterritorial boundaries often through internal ethnic passports, (2) quotas in the government for non-dominant groups or affirmative action, and (3) native language schools, holidays and theatres or cultural rights—all resourced by the state.

MRE was devised in the Soviet context to bring socialist order to ethnically disorganized populations (Stalin, 1913), rather than subsuming all cultures into one. The first-order intention of identity engineering in communist contexts was to incorporate illiterate people into the communist scheme by educating them through the nation-building stage of Marxist-Leninism in their native language en route to a cross-national socialist revolution. The multiethnic, minoritization aspect of MRE was part-and-parcel to nation-state building in the Soviet context. Stalin maintained that the national question for underdeveloped indigenous groups “can be solved only by drawing the belated nations and nationalities into the common stream of a higher culture” (Stalin, 1913, VI). Persuading people of the socialist agenda required building literacy in the case of illiterate people. For people literate in a minority language, it required concessions under the guise of cultural elevation within the national scheme. The multiethnic approach of the Soviet regime

Table 1: Multiethnic regime engineering

		<i>Family History</i>		
		() Majority	() Minority	() Mixed
<i>State designation</i>	(A) Majority	Identity recognition	Identity misrecognition	Identity nationalization
	(B) Minority	Identity misrecognition	Identity recognition	Identity Minoritization

was promoted as more humane than the homogenization approach of the West, but ultimately sought the ambitions of territorial and ideological control.

For (Phillips & Reus-Smit, 2019, 8), and my borrowing from their diversity regime term, engineering diversity is all about power: “order builders self-consciously organize and institutionalize diversity in ways that make cultural difference legible and controllable, and that reconcile the recognition claims connected to authorized forms of cultural difference with existing structures of power and privilege.” New political entities must connect the organization of diversity to the legislation of power—making management of “ethnic” identities the norm rather than the exception. MRE, in these same terms, carries an ethnopolitical goal to recognize diversity while ensuring that the cultural content of its ethnic forms serves the political goal of its regime.

Table 1 presents six assimilatory processes sparked by MRE based on first, whether the state has designated an individual’s family in the state-sanctioned majority (A) or minority group (B) and second, whether an individual’s family history is in the majority group (), minority group (), or a mix between the two (). **Majority group identity recognition** (A) is when the state deems individuals’ families who were already trending toward the majority identity as in the majority—validating their pre-existing identity. **Majority group identity misrecognition** is the reverse (A)—when the state incorrectly assigns a family or village as in the majority, despite having underlying minority groups features. **Minority group identity recognition** (B) is when the state gives minority rights to individuals who already were leaning in the direction of the minority ethnic group. From a national standpoint, this process demoted their identities—from a predominant ethnic group of a relatively undefined region to an incorporated minority of a larger, more tightly controlled union. However, at the local level, the shift from identity recognition was upward—often from rural peasant to a capacity to send all your children to school, attend socialist seminars yourself, and gain a new job in service to both your linguistic heritage and the communist party. This study focuses on the local level actors in these processes. **Minority group identity misrecognition** is the reverse (B)—when the state incorrectly deems a family with majority ethnic features is deemed to be in the minority. As minorities within a new minority, this population is the smallest group of the six. **Majority group identity homogenization** (A)—or identity nationalization—is when the state designates the uniform majority identity for a family whose members previously carried a mixed identity. **Minority group identity homogenization** (B)—or

minoritization—is when the state deems a mixed identity family to be in the minority group. Regardless of the pre-existing family histories, each of these processes are assimilatory because prior to communism, those targeted by MRE lived in an identity environment containing more than one choice of ethnicity. But in communist-engineered environments, the state reduced the ethnic choices to one—based on the territory in which you were born.

In brief, communist regimes were earnest in their efforts to recognize distinct minority groups. “Uncompromisingly hostile to individual rights, they eagerly, deliberately and quite consistently promoted group rights” (Slezkine, 1994, 415). These ambitions of a state or policy to promote group rights over and above the past identity claims of the population through ethno-territorial borders, affirmative action policies and education is the process which constitutes my primary independent variable. The case I use is communist-era southern Albania, which then, contained substantial Greek-speaking, Albanian-speaking, and bilingual populations. The impact of communist MRE on this population is the core motivation for this study.

2.2 Case Fit: Communist Albania as a case of MRE

Albania is a viable case study for tracking the identity transmission outcomes of MRE because Hoxha drew directly from Stalin’s playbook on minority identity recognition. He inscribed support for the Greek minority from the beginning of his reign (Hoxha, 1985, 262-63), and he kept his constitutional promise, on strict and measurable terms. He took a bifurcated approach which simultaneously resourced, on the one hand, the Albanian identity in rural populations; and on the other hand, recognized a territorially a designated Greek identity within Albania—both for the first time and exogenous to the previously mixed linguistic identities of the population. He mimicked the timeline of Stalin’s enforcement strategies over an expanded duration of time in a micro-context. Hoxha applied an ethnoterritorial border around the “true” Greeks and resourced the “Greek minority” identity with state provisions for (1) positions for Greeks in the military and government, (2) cultural rights, such as road signs, newspapers, and events in the Greek language, and (3) extensive educational programs, courses, and schools in the Greek language (Partia e Punës e Shqipërisë, 1953).

2.3 Sampling Frame: Treatment and Control Group Descriptions

Hoxha’s experiment with international and internal borders allows three groups for this study’s analysis: A) those born into Hoxha’s recognized Greek minority zone, B) those born outside the zone but in the same Albanian districts bordering Greece, and C) those who did not experience Hoxha’s MRE because they were born across the newly enforced international border with northern Greece. Groups A and B are found in within the districts of Gjirokastrë, Korçë, and Vlorë in today’s Albania. Group C is found in Thesprotia, Ioannina, Kastoria, and Western Macedonia within the Epirus district in today’s Greece.

There was little to no mobility between groups A/B (in Albania) and C (in Greece) after 1945 due to Hoxha's sweeping, sudden, and initial enforcement of the Albania-Greece border in 1945. Mass emigration ensued after 1990. Successful escapees between 1945-90 were around 800 and carefully documented by Greece and in international news. Unsuccessful escapees were documented carefully within Albania. Little mobility occurred between groups A from B from group C due to an internal ethnoterritorial border enforced by Hoxha after 1945. Cross-mixing, or migration between group A and group B was carefully documented and almost never occurred in line with the larger trend of rural-to-urban migration. Out-migration in general was carefully controlled and documented by the state. In-migration was also extremely uncommon. The pre-existing family histories are derived from the same population group, 1930-45.



Figure 1: Geographic Scope of Study

Group A received resources from the Albanian state for Greek language schools, holidays, and newspapers. Group B serves my analysis as a secondary treatment group. They received Albanian language resources, typically for the first time, regardless of their ethnolinguistic preferences. I break groups A and B into 3 sub-groups based on family history (Albanian/Greek/mixed). Greece-based group 3, the “control” group (C in Figures 1-2), are those who were cut off from Albania during the Hoxha period, receiving neither treatment.

Regarding timeline, I focus on those born to parents who experienced communist Albania after Hoxha came to power in 1945, focusing initially on those born between 1946-1965. Most adults born between 1946-1965 in a southern Albanian or northern Greek district neighboring the Albania-Greece border are

eligible for this study’s analysis—including those who died, escaped, or out-migrated after the Albania-Greece border became operative again in 1990. This means the target population for the first wave of this study is aged 57-76 (in 2022). For pre-period comparison, I also collect information on people born 1930-45 in the same locations of my study. Members of this group are aged 77-92 (in 2022).

2.4 Dependent Variable: Ethnopolitical Identity Compliance

Ethnopolitical identity compliance is when individuals get in line with their state-assigned ethnic and political identity content. “National in form, socialist in content,” was the mantra for a successfully engineered MRE state (Slezkine, 1994). To measure ethnic identity compliance, I use the number of people who adopted the state-assigned identity by choosing to name their kids from a pool of names linked to the assigned identity—drawn from censuses from before the pre-period of the study. The name is also counted as compliant if it is a new name if the parents broke from their family pool of first names, instead naming their child from a list of classical figures or recent national heroes. These lists of names were introduced through Hoxha’s nationalization program as superior to the traditional practice of naming children after grandparents and/or religious figures.

Long-term ethnic identity success is measured via stability of that state-assigned identity across one or more generations—marked by the number of people using the script associated with the state-assigned ethnolinguistic identity on their family’s grave. Long-term political identity success is measured by having raised children and grandchildren who affirm socialism today through beliefs expressed in favor of, and/or through votes for, Albania’s socialist parties. This is further gauged with a survey question capturing expressed admiration for and overlap with their most significant grandparent’s political views. Albania’s village cemeteries provide an important addition to identity markers recorded in the local registries because cemeteries were often physically located less in the purview of the state. Moreover, after the regime fell in 1990, many families added new headstones for people who died during Albania’s authoritarian period, 1960-90—thus, “correcting the record” for their parents and ancestors in a sense, whereas today it is not possible to update the state’s civil registry records in Albania. For this reason, Albania’s civil registry records offer the strongest measure of a parents’ identity trend (through the names of their children), and gravestones provide a better snapshot of the ethnic legacy memorialized by one’s children thereafter (through the script chose for the grave). Together, they capture short and long-term ethnopolitical identity compliance. Support for socialist platforms is a strong measure of political identity compliance in the post-communist period.

The definition for ethnopolitical identity compliance is agnostic to the sites or actors which shape an identity, paralleling the intuition of Fearon (1999)’s “identity-as-a-social-category” as well as Mylonas (2010)’s broad inclusion of religious and ethnic minorities under “non-dominant groups”, because in MRE contexts, identities could be prescribed by the state, stripped by the state, reassigned by the state, asserted

Table 2: Successful Ethnopolitical Identity Compliance in an MRE Context

		<i>Family History</i>		
		() Majority	() Minority	() Mixed
<i>State designation</i>	(A) Majority	A + socialist	A + socialist	A + socialist
	(B) Minority	B + socialist	B + socialist	B + socialist

by individuals, prescribed by local church authorities, and everywhere in between. Table 2 illustrates what ethnopolitical identity compliance should look like in a multiethnic state which is successful.

The secondary outcomes of this study constitute family and societal level outcomes targeted by the regime more broadly alongside the MRE allocation: tolerance for out-groups, cross-ethnic marriages, lack of female participation in the labor force, lack of participation in socialist courses, and atheist identifications. Success entails growth in these areas.

Each of these outcomes are measured as rates of change in proportions derived from individual-level statistics and logged at the village level. This is the same manner in which Hoxha’s party measured and evaluated the policy during communism¹—as well as the suggested setup for a DD design adjusting for a contrast between the groups in the pre-period (Angrist & Pischke, 2014, 130).

2.5 Pre-existing (Family Identity) Conditions

This study imputes data for three pre-existing identities measured at the family level. Because national identities were largely imposed on the region’s underlying diversity, a creative, comprehensive, and pre-specified approach is necessary for defining the identity direction of one’s family. Language is the primary component, but religious name banks are also used as a part of the measurement strategy because Orthodoxy was more closely associated with the Greek identity whereas Islam, as well as atheism, was more closely associated with Albanian identity. Moreover, most females were illiterate, meaning the civil registry record left their language column blank—making religious names a more feasible proxy for identity. To capture the trend of the family’s identity, I also include names assigned to children in the household.

Nationality was not collected in the 1930-45 civil registry. To account for this, I pre-specify definitions for family identity markers as a proxy for underlying trends. (1) I assign a person’s family identity as Greek if their family’s head of household was literate in Greek and all children in the household were given Orthodox names, 1930-44. (2) I record a person as Albanian if the head of household was literate in Albanian, married someone with a non-Orthodox name (whether Muslim or non-religious), and all children in the household were given non-Orthodox names. (3) They are deemed as carrying an identity

¹Success of the policy was gauged quantitatively with elections for the communist party, more party members joining, increases in electoral turnout, socialist course attendance and qualitatively, i.e., the authenticity of the course discussions about socialism, the degree to which community members interacted with foreign-based relatives, whether attempted escapes had decreased, whether or not people expressed admiration for or rejection of religion, and whether or not people had ostracized community members deemed “kulaks” by the communist party.

mixed between Greekness and Albanian-ness if the head of household married someone with a name of the opposite religious and/or ethnolinguistic group and they named their children with first names matching both pools of religious names (Orthodox/Muslim). (4) On the rare occasions people are listed without a specified family history, I record them as having no pre-existing conditions.

2.6 Theory

I test the theoretical argument in historical studies that MRE spurred the formation of minority ethnicities but failed politically. In support of this theory, [Roeder \(1991\)](#) found variation in outcomes of MRE: those with nationalities that had previously high levels of socioeconomic standing (i.e., the Armenian elites in the Caucasus) effectively unraveled the Soviet MRE policy to their benefit, whereas national elites in less developed areas (i.e., the Kazakhs in Central Asia) were less successful in pressing for sociopolitical agendas and empowering themselves against the Soviet identity [Gurr \(1993\)](#); [Laitin \(1998\)](#). [Roeder \(1991\)](#) calls this regional divide in outcome between national groups ironic since the idea behind MRE policies was to level out national groups around a central Soviet power—which the policy did succeed in doing, in terms of cohering lesser-renowned ethnic identities, but simultaneously reaffirmed the preexisting inequalities and political hierarchies among ethnic groups which communism was designed to overcome. In other words, MRE, achieved certain “ethnic” outcomes—for instance, growing written use of the Ukrainian language over time—but failed politically in the long-term. [Treisman \(1997\)](#) terms the political failures of MRE “ethnic revival,” affirming previous literature with an instrumentalist explanation: the relatively privileged minority groups leveraged their minority status against other groups and against the central power. In a similar vein, [Slezkine \(1994\)](#) and [Gorenburg \(2003\)](#) argue that the Soviet government implemented a dual course for its minority groups. Albania’s communist party memos supports this literature—with evidence that Hoxha both actively resourced the Greek minority identity with expansive translation assignments and that those translation efforts were highly targeted toward transmission of the Soviet socialist ideology, ([Partia e Punës e Shqipërisë, 1953](#)). These memos allow this study to specify precisely what Hoxha considered failures of the policy.

Beyond post-Soviet studies, the case of Albania allows me to test overarching claims made within the literature about the transmission process of ethnic groups during and after communism. For instance, with census data, ([Gorenburg, 2003](#)) illustrates the byproducts of MRE for minority groups within territorial Russia: they seemingly chose to assimilate as Russians. However, the process in which they did so is unclear, and whether those shifts in identity stuck in the long-term is ambiguous. He makes a further, untested claim: “Had the Soviet Union retained such policies for another 1-2 generations, it is not unlikely that a large percentage of minority group members would have declared Russian as their native language or switched their ethnic identity to Russian” ([Gorenburg, 2003, 27](#)). A similar finding in my context will be if Greeks who fell outside the zone, over-time, became Albanians. This untested claim is interesting

in its own right—and the core substantive motivation for this study.

Rather than contesting post-Soviet literature, I refine and expand the theory that MRE forged ethnicities and failed politically with an original micro-level dataset curated from the districts surrounding Albania-Greece border. The micro-context of southern Albania affords the opportunity to detail the conditions under which state-led assimilation succeeds. My overarching theory is that the extent to which Multiethnic regime engineering succeeded within individuals corresponds to two factors: 1) whether they were born into a village recognized as a minority within the regime’s ethnoterritorial border and 2) their family’s pre-communist level of identity attachment toward one cultural identity or the other (measured with a family-level score composed of linguistic and religious features, as well as parental origins). Specifically, I predict that, on average, Hoxha’s ethnic identity designations in southern Albania largely succeeded, and when they failed, it was due to pre-existing identity conditions of families who lived outside the recognized contiguous linguistic zone. The reasoning behind my hypotheses is the intuition that MRE succeeded ethnically under conditions in which people had the opportunity to obtain a new identity from previously mixed identities but failed politically where they were forced to assimilate into a majority identity from a previously strong minority identity background. As such, I posit two overarching, long-term outcomes for the groups bifurcated in an MRE environment.

- A. Those who did not receive the minority designation, despite previously having greater access and exposure to the identity, became more likely to ardently adhere to either identity—based on their prior family histories. Overall, this group was still likely to pass on the ethnic identity assigned by the state but fractured politically in the long-term along the lines of past family and local-level identities. Regardless of whether they have previous attachment to either group, I expect they were overall more likely carry sectarian attitudes and behaviors at odds with the ideals of socialism relative to the recognized group.
- B. Those assigned a privileged minority group became likely to tow the communist line, thus, passing on the assigned ethnic identity to their kids as well as conformist attitudes in relation to the local governmental power and socialist attitudes.

3 Specified Hypotheses

My directional hypotheses are as follows:

1. **Majority group identity recognition:** I predict that those assigned a majority identity (A), while carrying identity features toward the majority group (), complied with the communist regime’s majority identity designation during communism and leaned in the direction of communist party platforms during and after communism (A) A), but were more likely to adopt long-term nationalist, sectarian political identities. I posit initial ethnic success of this group because their

identity assignment entailed a good fit and the most gains in Albania’s early communist period, whereas I expect to observe political failure of the policy in the long-term because they suffered during and especially after communism from the relative privilege of the minority in international settings—leading to long-term resentment and higher levels of sectarian attitudes post-communism.

2. **Majority group identity misrecognition:** Those subsumed into the majority designation (A), despite coming from a family with minority identity features (A), I predict are the most likely group to show ethnic identity failure—having transmitted identities with pre-communist identity features (), aligning more with the general direction of names in northern Greece than in the recognized minority zone. I also expect they were most likely to fail to comply with the regime during communism and that they leaned away from socialism in the direction of rightwing, anti-socialism party platforms post-communism.
3. **Majority group identity homogenization:** Those assimilated into the majority group, while carrying pre-communist transnational identities (), I expect leaned toward their assigned ethnonational identity () A) and toward socialist parties post-communism.
4. **Minority group identity misrecognition:** Those designated for the minority group (B), despite carrying opposite features of the majority group (), I expect adopted the state-assigned minority identity () B) and trended in favor of the communist political identity during communism, being more likely to carry socialist attitudes post-communism.
5. **Minority group identity recognition:** Those who received a minority designation within the regime’s territory (B), carrying features in the qualitative direction of the recognized identity (), I also expect adopted the assigned linguistic identity () B), but with measurable differences in name use trends from those across the foreign border unaffected by the policy—demarcating a “minority identity” resulting from MRE. I also expect they trended toward support for the communist party during communism and toward centrist political party platforms post-communism.
6. **Minority group identity homogenization:** Those nationalized into the minority group (B), despite previously carrying transnational identities (), I expect leaned into the identity of the minority group () B) and adopted socialism attitudes.

The study’s second line of outcomes are types of failures to comply with MRE. I predict that those not recognized as a minority, groups (A) and (A), will carry higher levels of intolerance for expressions of each other’s out-group identity, compared to groups (a) and (b). I expect group (A) to be most likely to carry an atheist identity, during and after communism—whereas groups (B), (B), and (A) will be more likely to shift toward religious identities post-communism. The strongest within-family

Table 3: Hypotheses Overview for Ethnopolitical Identity Compliance

		<i>Family History</i>		
		() Majority	() Minority	() Mixed
<i>State designation</i>	(A)Majority	Recognition) <i>A</i> socialism	Misrecognition) anti-socialism	Homogenization) <i>A</i> socialism
	(B)Minority	Misrecognition) <i>B</i> socialism	Recognition) <i>B</i> socialism	Homogenization) <i>B</i> socialism

identity continuity I expect will be within the minority zone, groups (B) and (B). I expect the highest out-migration rates to be from group (A).

4 Relevant Background and Scope of Study

4.1 Literature

The field of post-Soviet studies contains a robust body of research which seeks to understand the legacies of communist regimes on institutions, attitudes, behavior, and identity in transnational contexts (Acemoglu, Johnson, & Robinson, 2002; Bernhard & Karakoç, 2007; Crawford & Lijphart, 1995; Darden & Grzymala-Busse, 2006; Hagopian, 1993; Laitin, 1998; Neundorf, 2010). However, a lack of reliable intergenerational data has meant that the long-term effects of communist-era identity engineering are under-specified (Pop-Eleches & Tucker, 2020). Moreover, also due to data difficulties, the mechanics of identity transmission in broader literature is most often assumed rather than measured, i.e., in anthropology (Rogers, 1988), game theory (Bednar & Page, 2007), economics (Alesina, Giuliano, & Nunn, 2013; Algan & Cahuc, 2010; Michalopoulos, 2012; Nunn & Wantchekon, 2011; Voigtländer & Voth, 2012), and political science (Acharya, Blackwell, & Sen, 2016). To address this empirical gap, I draw from the micro context of Albania, given the duration for and consistency with which its minority identity engineering policy was applied during communism, 1945-65—and the recent availability of micro-data from this period’s population.

The literature available on the Greek minority in Albania are ethnographies (Bon, 2008; de Rapper, 2004; Nitsiakos, 2010; Schwander-Siever & Fischer, 2002), histories (Binder, 1984; Giakoumis, 2003; Giakoumis & Kalemaj, 2017; Howe, 1983; King & Vullnetari, 2012; (Papa), 2011; Pettifer, 2000; Pollo & Puto, 1981; Stickney, 1926; *Thirty-three Albanians Flee by Boat to Greek Island of Corfu*, 1971), linguistic studies (Kyriazis & Qiriazzi, 2007), journalistic accounts (Howe, 1983; Stokes, 1987), and first-person narratives of foreign diplomats who wrote about their experiences in communist or post-communist Albania (Abrahams, 2016; Jones, 1994; Keeley, 1997). These accounts give me a strong sense the Greek minority in Albania, and those excluded, have an untold political story. From the empirical claims of this

literature, as well as a review I conducted at Albania's Ministry of Interior, I stage the relevant history for this study.

4.2 Hoxha's Experiment

The process of building the Albanian national state began later than the longer-standing nation-state projects in the region, such as Greece. Albania was the last of the Ottoman states to declare independence in 1912 and was ruled by King Zog between 1922-39. In this period, what is today's southern Albania composed a disputed territory between Albania and Greece called "Northern Epirus" (*Vorios Ipiros*). Vórios Ípiros was the central point of more than 10 international border commissions tasked with measuring ethnicity prior to 1930. A United States (U.S.) funded dissertation on the topic concluded from these records that the ethnographic distinctions of this region were impossible to define, and the population's cultures, religions and languages were mixed (Stickney, 1926, 3).² My review of the Albanian civil registry taken soon after Stickney's study concludes similarly. Albania's civil registry records, 1930-45, indicate fluidity of identities in this region prior to 1945, between Albanian and Greek ethnic features as well as between Muslim and Orthodox religions—with general trends toward Orthodoxy as the higher status religion. The category of ethnicity was not surveyed in southern Albania before Hoxha's 1945 national census. Instead, scholars and order commissions had documented religious identifications and/or language(s) spoken.

After Hoxha took power in 1945, he launched the process of ethnic engineering in explicit imitation of Stalin's approach in the Soviet Union. More forcefully than Stalin, Hoxha halted movement in/out of Albania, separating families for decades. Hoxha's 1945 census,³ is suggestive of calculated "identity switching" occurring at the village level between Greek and Albanian identities when paired against the 1930-45 civil registry. These statistics, however, are unreliable on their own and do not align with Hoxha's ethnoterritorial border-making. Hoxha drew minority zone boundaries based on two inclusion criteria, one of which was old and one of which was new. First, the village had to be one which had already gained formal rights to open a school in the Greek language during the previous monarchical period in Albania under King Zog. Second, he modified that the village had to fall within a contiguous zone at the foreign border with Greece. This meant that villages which previously had formal Greek rights and/or access to significant Greek influence, were excluded—instead, receiving new Albanian language schools and general assimilation program toward an Albanian national identity. Minority groups without previous territorial rights or a foreign border—such as Vlachs, Cham Muslims, Bektashi Muslims, Romanians, and Roma people—were disregarded by the state. Hoxha deemed use of the term *Vorios Ipiros* traitorous, and

²The population was largely bilingual. \Nowhere else in Europe is the criterion of spoken language so fallacious as in the Balkan states." (Stickney, 1926, 3). \An observer of Albanian sympathy might pass through the country-side and find nothing but Albanians, where an observer of Greek sympathies proceeding in the same manner might find nothing but Greeks" (Young in (Stickney, 1926, 3)).

³Hoxha's census took a remarkably similar campaign-style approach to the Soviet Union's census-taking and border-making approach detailed in Hirsch (2014).

socialism very quickly became the only allowable political game in town. Those born in Albania in 1946 or after were afforded state provisions for Greek, Albanian, or Macedonian status—based strictly on the territory in which they were born. These rights did not travel outside of the minority zones. However, one could marry into the minority zone, affording rights to one’s children, as well as attend celebrations in the minority zone, as did Hoxha on several occasions. Regardless of one’s past, all people born in southern Albania after 1945 experienced a more territorially stagnant identity environment than did their parents and grandparents.

Those who lived in neighboring districts of today’s northern Greece did not experience the Multiethnic regime engineering policy. Instead, they experienced the homogenizing approach of the Greek state toward a unified Greek national identity, centered around attachments to 1) the Greek language, 2) the Orthodox religion, and 3) the state’s narrative of Greek territory. Hoxha’s minority identity designations continued from 1946 through 1985, the year he died, without the option of exit. After 1990, when the communist regime collapsed, Hoxha’s Greek minority zone continue to be recognized by the Albanian and Greek states with greatly increased opportunities for exit. Since 1990, more than half the population out-migrated. These historical conditions set the stage for my study.

The timeline of 1945-65 of Albania’s communist period is useful for analysis because it was prior to Albania’s Mao-inspired “Cultural and Ideological Revolution” (1966), when Hoxha’s identity engineering veered sharply toward ethnically indiscriminate repression of Albania’s entire religious population. This allows me to specify the treatments of the study to be a lighter form of authoritarianism and in fact, include incentives to buy-in to the communist ethnic designation, such as schools, libraries, hospitals and the like. Moreover, focusing on the children of those born into communism, rather than their parents who experienced it initially, induces additional, if limited, randomization to the study’s population. Additionally, focusing on the children whose parents encountered Hoxha’s communist takeover also helps account for the fact that, prior to Hoxha’s 1946 border closure, ardent supporters of the rightwing “Northern Epirus” movement had the opportunity to exit.⁴ For these reasons, I focus especially on those born 1945-65. However, in a second wave of this project, I will extend it to Hoxha’s more culturally repressive period and update my specified hypotheses accordingly.

5 Design

My DD design leverages Hoxha’s sudden, sweeping closure of Albania’s borders, which separated current and future generations’ access to each other and bifurcated my study’s target population at the Albania-Greece border into communist and non-communist lived experiences respectively. Prior to 1945, the

⁴These political departures were rare, however, and measurable since Hoxha kept close track of Albanian families with a family member who fled. Most were men planning either to return or to move their family with them once they had resettled successfully in Athens, Greece. However, they were deemed enemREs of the state by Hoxha and were unable to return, just as those in Albania were unable to visit them due to the 1946 border closure. With access to these previously sealed records, I can account for this influence in this study.



Figure 2: "Toast to Brotherhood," Painting of an Albanian hero, Çerçiz Topulli, linking arms with a Greek, surrounded by his çeta, a multiethnic military band who legendarily fought off the Ottomans together. Hoxha added this image with the following statement in his 450-page book on the Greek minority: "The Greek minority and the Albanian people have always been linked as closely as flesh to bone, at all times they have loved and supported one another."

Albania-Greece border was unmarked and practically meaningless to the local population (de Rapper, 2004; Schwander-Siever & Fischer, 2002). Hoxha’s border enforcement thereafter separated families and halted free movement for decades (between 1945-1990). I also leverage for two treatments in the analysis Hoxha’s insistence that the political rights of Albania’s Greek minority be recognized only in a contiguous territory, “bordering a foreign state”—a criterion drawn from the MRE approach taken in the Soviet Union (Hirsch, 2014). In practice, this meant that only villages with 100 percent of the heads of households identifying as Greek speakers *and* living in a contiguous territory became part of the approved “minority zone.” This was even though Greek speakers and people with literacy in Greek—as well as Orthodox identifiers—lived throughout the south of Albania. This international border closure and simultaneous internal barrier on ethnicity construction created three territorially based groups for analysis: 1) those deemed Greeks in Albania, 2) those deemed Albanians, and 3) those who were born in what became modern-day Greece, experiencing neither designation but rather, the homogenization approach of the Greek state.

5.1 Model

Y_{vt} refers to the share of people who complied with their assigned ethnopolitical identity, where subscript v illustrates whether the data is from a village in the group deemed at the majority/minority (A/B) and subscript t denotes whether the data is for names assigned before 1945 (in the civil registry taken 1930-44, before Hoxha’s MRE application) or 1945-1960 (after). The DD estimate ($DD_{;A}$) of the effect of nationalizing a diverse population into the Albanian majority group is:

$$DD_{;A} = (Y_{A;1945} - Y_{A;1930}) - (Y_{B;1945} - Y_{B;1930}).$$

The DD estimate ($DD_{;B}$) of the effect of recognizing the Greek minority in Albania swaps these terms or places a negative sign in front of ($DD_{;A}$).

$$DD_{;B} = (Y_{B;1945} - Y_{B;1930}) - (Y_{A;1945} - Y_{A;1930})$$

To make the comparisons more precise, I add a control group from northern Greece, which allows me to apply a type of “difference in difference in differences” (DDD) model, comparing comparable groups before and after Hoxha’s communist takeover, with two separate equations:

$$DD_{;A} = (Y_{C;1945} - Y_{C;1930}) - (Y_{A;1945} - Y_{A;1930})$$

and

$$DD_{;B} = (Y_{C;1945} - Y_{C;1930}) - (Y_{A;1945} - Y_{A;1930}).$$

I further refine the analysis to examine the impacts by sub-group by matching comparisons based on their pre-existing family histories. Thus, the DD estimate of recognizing an Albanian majority family correctly with their prior family trend, is $DD_{;A} = (Y_{C;1945; } - Y_{C;1930; }) - (Y_{A;1945; } - Y_{A;1930; })$. For the estimated effect of misrecognition of a Greek family as Albanian: $DD_{;A} = (Y_{C;1945; } - Y_{C;1930; }) - (Y_{A;1945; } - Y_{A;1930; })$. For the effect of recognition of a Greek trending family as in the Greek minority

of Albania: $DD:B = (Y_{C;1945; } - Y_{C;1930; }) - (Y_{B;1945; } - Y_{B;1930; })$. For the impact of misrecognition of an Albanian trending family as Greek: $DD:B = (Y_{C;1945; } - Y_{C;1930; }) - (Y_{B;1945; } - Y_{B;1930; })$. For the effect of assimilation into the majority on mixed families: $DD:A = (Y_{C;1945; } - Y_{C;1930; }) - (Y_{A;1945; } - Y_{A;1930; })$. For assimilation into the minority for mixed families: $DD:B = (Y_{C;1945; } - Y_{C;1930; }) - (Y_{B;1945; } - Y_{B;1930; })$.

5.2 Parallel Trends

A DD design makes the assumption that without the intervention of the treatment group(s), the trends would have remained in parallel—thus, also resting on the assumption that the pre-period trends were initially parallel. Here I illustrate the parallel trends of the pre-period data in relation to ethnopolitical identity. **[**Add data here**]**

5.3 Sample size and selection

I draw my sample comprehensively from 26+ villages, yielding 7800 possible observations.

The predicted sample size across the 6 sub-groups.

[Add data here**]**

6 Dataset Description Overview

6.1 1930-45 Civil Registry Records

Dataset Curators With an Albanian-language research assistant (RA), Kristian Peçi, I hand-transcribed for this study all data available via *Familysearch.org*. The data was hand-written and more extensive than a traditional census. It included identifying features from religious identity and linguistic capacity to facial characteristics for each individual and was updated over time. Both historians and the Albanian government refer to these hand-written books as Albania’s first civil registry. Now this portion is available in a machine-readable format, i.e., an Excel spreadsheet.

Dataset Columns Albania’s pre-communist civil registry columns include village name, link to scanned version of the document per household, household ID, first name, surname, father’s first name, mother’s first name, gender, relation to head of household, birth year, birthplace, civil status, literacy, occupation, religion, death year, village the individual migrated to or from, and the year of in/out migration. A full description of each variable as well as brief contextual notes about the transcription process or data quality are provided here. In Appendix A, I expound the dataset variables in full. In Appendix B, I report an exhaustive list of village names/spellings for the village data transcribed at time of writing. I have not yet analyzed this data.

Creation dates The data was curated between 1930 to 2022. My collection of the data began in 2019 and the transcription process continues at time of writing.

Languages used in the data Albanian, Greek, and English

6.2 Cemetery data Overview

Data set creators In October 2021 and March 2022, I collected data from public cemeteries in Albania by taking photos of each grave—using the assistance of RAs Urania Gjoci, Ilia Gjoci, Bujar Ago, Sara Ago and one who wishes to remain anonymous. Greek and Albanian-speaking RA, Igri Mama, Sara Ago and I transcribed the photos using pre-set definitions for each category of interest. The data currently spans the south of Albania, 1800 to 2022. I employed RA assistance from one RA who wishes to remain anonymous as well as local historian and tour guide, Bujar Ago. RA Sara Ago assisted in the field and transcribed a substantial portion of the data. The data collection and transcription processes are ongoing.

Data set Columns Each row of the dataset corresponds with a memorial devoted to an individual and includes: a link to a photo of the grave, the village name, district, geolocation, site type, script of written info, gender, last name, first name, birth year, religious symbols, political symbols, eulogies, translation of the eulogy (when relevant), and whether or not someone left an identity marker at the grave recently. In Appendix C, I explain the memorial data I have collected throughout southern Albania and transcribed into a machine-readable format. The completed digitization process includes 47,374 observations derived from 212 villages.

Creation dates The earliest legible death date in the cemetery data is 1876, and the graves span to 2021.

Languages used in the data Albanian, Greek, and English

6.3 Addressing Weaknesses

Unobservables With any causal inference design using observational data, unobserved confounding variables are a possibility ([Jake Bowers & Ichino, 2022](#)). To address this, I chose a micro-context with extensive data available on the individual, family, village, and state levels—within a relatively closed setting. I also conducted a thorough review of available archival records and consulted several experts and local historians who studied and/or experienced the policies of the regime. Regarding unseen background conditions, my study’s geographic scope previously was part of an Ottoman-era vilayet, Janina. Vilayets had no internal specified administrative structures—reducing chances of unseen background conditions related to the international border Hoxha enforced between the populations. Moreover, the information

of the civil registry combined with public archives allows me to build the following potential covariates into my model: income, gender, previous migration experience, communist party member, "kulak" status, and networks with family in the United States and/or Greece documented by the communist party.

Spillover as Byproduct Effect Regarding spill-over between the two groups 1 and 2, this is to be expected since intermarriage between the two groups was practiced prior to communism and was permitted by Hoxha. However, only those born inside the minority zone received Greek minority status. To account for spill-over, I track intermarriages between the groups and population disparity as secondary outcomes of the study. Regarding spill-over between groups 1-2 and 3, this is also to be expected given the brutality of Hoxha's regime. However, exiting was not permitted and both unsuccessful and successful attempts were carefully counted, documents recently made available to the public. As such, I am interested from the outset of the study in the extent of the spillover between treated groups—as a byproduct of the policy.

Biases and Missing Data Deriving outcome data from state records for a policy enforced by the same authoritarian state is inadequate on its own. The grave data is also incomplete on its own since roughly half of the affected population is still alive and/or has migrated permanently.⁵ For a more complete picture, I use both, which allows me to compare representations of people's identities to the state with representations of one's family to their local communities for the same subset of individuals.

Theoretical Issues with Name Data The names given children is a strong measure of ethnopolitical identity in Albania. Prior to communism, most people in the south chose names for children based on the religion to which they subscribed. Moreover, in the Orthodox tradition, names are officially given by the child's godparent, although they were most likely chosen in consultation with the parents and traditionally, would correspond with grandparent's names. As such, prior to communism, first names were primarily a social decision made within one's religiously affiliated network. During Albania's communist period, first names remained a decision made within social spheres, but in 1960 a national law was passed, requiring people with religious names to change their names. For this reason, I supplement state records with cemetery records, where I expect many people have been memorialized in the name their parent's initially gave them—rather than the ones forced to be changed in state records. The cemeteries also allow a recognizable measure of scripts chosen by families.

To supplement the name data, I also use other columns in the civil registry record, such as the literacy columns, and nationality column added during the communist regime. Moreover, choosing a name like "Enver" or "Stalin" is a strong measure of ethnopolitical identity compliance, but *not* choosing a name in compliance with the state identity should not be interpreted as politically motivated lack of identity

⁵The relatives of more than 6,000 people have reported a family member as missing to the International Commission of Missing Persons.

compliance. To supplement the political dimension for failure to comply, I collect political identity symbols at graves, such as communist stars, as well as behaviors, such as vote shares for socialist parties.

6.4 Linking the Data

My data collection approach allows me to link markers of identity for the same individuals over their lifetime as well as from the same families across four generations. Moreover, I use two sources of data for the same individuals, one recorded by local state officials and one publicized by families after death.

[**Add brief overview of linking process**]

7 Discussion

8 Conclusion

This study sets out to examine ethnic and political identity transmission before, during, and after communist-era identity engineering policies—focusing on the long-term impacts of multiethnic policies applied during communism in Albania. Research on communist-era Albania, one of the world’s most brutal authoritarian regimes, requires a mixed methods approach, creative measurement strategies and experience. [Pop-Eleches and Tucker \(2020\)](#) conclude in their review of legacy research on communist countries that the unanswered question of identity outcomes in post-communist countries should be tested with data better suited for testing cross-generational transmission processes. I use individual level behavioral outcomes, designed to capture ethnic and political identities as the primary outcomes of the study. I capture religious identities as a secondary outcome of the study. I also record each village’s public monuments, memorials, churches, and mosques at each site. While not a primary or secondary outcome of interest, this monument data offer insight into whether the current village identity entrepreneurs hearken toward the communist past. With a hypothesis Multiethnic regime engineering contributed to ethnic identity compliance but failed politically, I expect to find that overall, ethnic identities diverged during communism, in line with Hoxha’s rules. But those who were misrecognized after their families had been trending toward a specific ethnic identity, I expect became the most likely to maintain anti-socialist attitudes which persist today among their children. [**Add results**] With micro-level data coverage of a country’s entire national period, 1913-now, this study affords us one of the closest looks at the relationship between authoritarian states and the process of building as well as transmitting identity. Future research can benefit from the data—especially for the question of the costs versus benefits of complying with versus defying a state’s identity designation for future generations.

Acknowledgements

I acknowledge outstanding research assistance from Kristian Peçi, Igri Mama, Sara Ago, Bujar Ago, Urania Gjoci, and Ilia Gjoci—each of whom play a critical role in collecting and/or processing the data for this paper. I am especially grateful for research supervision and guidance from my dissertation committee members in political science, Susanne Wengle, Victoria Hui, Jaimie Bleck and David Laitin—and from historian Konstantinos Giakoumis. I am grateful also for my many informal readers of this paper in various stages—especially Vassiliki Fouka and Nicholas Haas for preparing discussant comments for the paper’s improvement.

Funding Statement

The data digitized for this research resulted from an American Political Science Association Doctoral Dissertation Improvement Grant (APSA DDRIG). I also acknowledge support from APSA and the Kellogg Institute for International Studies for funds to present early versions of this project at the 2022 Midwest Political Science Association (MPSA) conference in Chicago and at the 2022 European Political Science Association (EPSA) in Prague. Additionally, I acknowledge research awards from the American Institute for Southeast European Studies (AISEES) foundation and from the Association for Slavic, East European, and Eurasian Studies (ASEEES).

Competing interests

I declare no competing interests.

References

- Abrahams, F. C. (2016). *Modern Albania: From Dictatorship to Democracy in Europe* (Reprint ed.). New York University Press. DOI: <http://dx.doi.org/10.18574/nyu/9780814705117.001.0001>
- Acemoglu, D., Johnson, S., & Robinson, J. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics*, 117, 1231-1294. DOI: <http://dx.doi.org/10.1162/003355302320935025>
- Acharya, A., Blackwell, M., & Sen, M. (2016). The Political Legacy of American Slavery. *Journal of Politics*, 78, 621-641. DOI: <http://dx.doi.org/10.1086/686631>
- Alesina, A., Giuliano, P., & Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *Quarterly Journal of Economics*, 128, 469-530. DOI: <http://dx.doi.org/10.1093/qje/qjt005>
- Algan, Y., & Cahuc, P. (2010). Inherited Trust and Growth. *American Economic Review*, 110, 2060-2092. DOI: <http://dx.doi.org/10.1257/aer.100.5.2060>
- Angrist, J., & Pischke, J.-S. (2014). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.
- Bednar, J., & Page, S. (2007, 2). Can Game(s) Theory Explain Culture? *Rationality and Society*, 19, 65-97. DOI: 10.1177/1043463107075108
- Bernhard, M., & Karakoç, E. (2007, 7). Civil Society and the Legacies of Dictatorship. *World Politics*, 59, 539-567. DOI: 10.1353/wp.2008.0001
- Binder, D. (1984, 11). *Albania Opens its Doors a Crack*. Retrieved from <https://www.nytimes.com/1984/11/11/magazine/albania-opens-its-door-a-crack.html>

- Bon, G. (2008). *Contested Spaces and Negotiated Identities in Dhermi/Drimades of Himare/Himara Area, Southern Albania*. Retrieved from <https://www.ung.si/-/library/doktorati/interkulturni/3GregoricBon.pdf>
- Crawford, B., & Lijphart, A. (1995). Explaining Political and Economic Change in Post-Communist Eastern Europe. *Comparative political studies*, 28, 171-199. DOI: 10.1177/0010414095028002001
- Ş. Aktürk. (2011). Regimes of Ethnicity: Comparative Analysis of Germany, the Soviet Union/Post-Soviet Russia, and Turkey. *World Politics*, 63, 115-115. DOI: <http://dx.doi.org/10.1017/S0043887110000304>
- Darden, K., & Grzymala-Busse, A. (2006, 10). The Great Divide: Literacy, Nationalism, and the Communist Collapse. *World Politics*, 59, 83-115. DOI: 10.1353/wp.2007.0015
- de Rapper, G. (2004). 'We Are Not Greek, but...': Dealing with the Greek-Albanian border among the Albanian-speaking Christians of Southern Albania. *Southeast European and Black Sea Studies*, 4, 162-174. DOI: 10.1080/14683850412331321778
- Fearon, J. D. (1999). *What is Identity (As We Now Use the Word)?* Retrieved from <https://web.stanford.edu/group/fearon-research/cgi-bin/wordpress/wp-content/uploads/2013/10/What-is-Identity-as-we-now-use-the-word-.pdf>
- Giakoumis, K. (2003, 1). Fourteenth-century Albanian Migration and the 'Relative autochthony' of the Albanians in Epeiros. The case of Gjirokastër. *Byzantine and Modern Greek Studies*, 27, 171-183. DOI: 10.1179/byz.2003.27.1.171
- Giakoumis, K., & Kalemaj, I. (2017). *Fortifying the Nation: The Image of the Greeks in Albanian History Textbooks (1945-1990)*. VR unipress. DOI: 10.14220/9783737008112.195
- Gorenburg, D. (2003). *Minority Ethnic Mobilization in the Russian Federation*. Cambridge University Press.
- Gurr, T. (1993). *Minorities at Risk: A Global View of Ethnopolitical Conflicts*. United States Institute of Peace.
- Hagopian, F. (1993, 6). After Regime Change: Authoritarian Legacies, Political Representation, and the Democratic Future of South America. *World Politics*, 45. DOI: 10.2307/2950726
- Hirsch, F. (2014). *Empire of Nations: Ethnographic Knowledge and the Making of the Soviet Union*. Cornell University Press.
- Howe, M. (1983, 10). *Something Seems to Stir inside Isolated Albania*. Retrieved from <https://timesmachine.nytimes.com/timesmachine/1983/10/23/248688.html?pageNumber=21&auth=login-smartlock>
- Hoxha, E. (1985). *Two Friendly Peoples: Excerpts from the Political Diary and Other Documents on Albanian-Greek Relations 1941-1984*. Tirana: "8 Nentori" Publishing House.
- Jake Bowers, M. V., & Ichino, N. (2022). *The Theory and Practice of Field Experiments: An Introduction from the EGAP Learning Days*. Evidence in Governance and Politics, EGAP.
- Jones, L. (1994). *Biogra: A Traveller's Tale*. Harcourt Brace.
- Kaser, K. (2014). *Family and Kinship in Albania: Continuities and Discontinuities in Turbulent Times* (Vol. 15). Lit.
- Keeley, E. (1997). *An Albanian Journal: The Road to Elbasan*. White Pine Press.
- King, R., & Vullnetari, J. (2012, 7). A Population on the Move: Migration and Gender Relations in Albania. *Cambridge Journal of Regions, Economy and Society*, 5, 207-220. DOI: 10.1093/cjres/rss004
- Kyriazis, D., & Qiriazi, D. (2007). [The relationship of the Greek Linguistic Idiom of Himara with Other Modern Greek Idioms]. Proceedings of the Annual Meeting of the Department of Linguistics, Aristotle University of Thessaloniki.
- Laitin, D. (1998). *Identity in Formation: The Russian-Speaking Populations in the Near Abroad*. Cornell University Press.
- Martin, T. (1998). The Origins of Soviet Ethnic Cleansing. *Journal of Modern History* 70, no. 4, 813-861. Retrieved from <http://nrs.harvard.edu/urn-3:HUL.InstRepos:3229636>
- Martin, T. (2001). *The Armative Action Empire*. Cornell University Press. DOI: 10.7591/9781501713323
- Michalopoulos, S. (2012, 6). The Origins of Ethnolinguistic Diversity. *American Economic Review*, 102, 1508-1539. DOI: 10.1257/aer.102.4.1508
- Mylonas, H. (2010). *Rethinking Violence* (E. Chenoweth & A. Lawrence, Eds.). The MIT Press. DOI: 10.7551/mitpress/9780262014205.001.0001
- Neundorf, A. (2010, 10). Democracy in Transition: A Micro Perspective on System Change in Post-Socialist Societies. *The Journal of Politics*, 72, 1096-1108. DOI: 10.1017/S0022381610000551
- Nitsiakos, V. (2010). *On the Border: Transborder Mobility, Ethnic Groups and Boundaries Along the*

- Albanian-Greek Frontier*. LIT Verlag.
- Nunn, N., & Wantchekon, L. (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review*, 101, 3221-3252. DOI: 10.1257/aer.101.7.3221
- (Papa), E. P. (2011). Religion and Gender in Albania 1967-2009. *Occasional Papers on Religion in Eastern Europe*, 31, 19-23.
- Partia e Punës e Shqipërisë, P. (1953, 14/APSTR). Report [...] në rrethin e Sarandës [...] mbi gjendjen dhe punën politike në zonën e minoritetit dhe të kufirit (Report [...] in the District of Saranda [...] on the Situation and Political Work in the Minority and Border Area). *Komiteti Qendror i Partisë së Punës së Shqipërisë - Strukturat e Komitetit Qendror (Central Committee of the Party of Labor of Albania - Structures of the Central Committee)*, 1-20. Retrieved from <https://katalogu.arki.va.gov.al/publik/categoris/438704-raport-e-relacion-te-instruktove-te-aparati-t-te-komiteti-t-qendror-te-partise-se-punes-se-shqiperise-dhe-ekipe-per-problemet-qe-kane-pare-e-ndjekur-ne-rrethin-e-sarandes-si-propaganda-e-partise-per-kolktivizimin-e-bujqesise-mbi-gjendjen-dhe-punen-politike-ne-zonen-e-minoriteti-dhe-te-kufirit-mbi-kualitetin-e-punes-ne-format-e-arsimit-te-partise-e-te-riunise-mbi-arsimin-shteteror-mbi-punen-politike-te-zhvilluar-ne-rrethin-e-sarandes>
- Pettifer, J. (2000). The Greek Minority in Albania: Ethnic politics in a Pre-National State. In J. Stein (Ed.), *The Politics of National Minority Participation in Post-Communist Societies: State-Building, Democracy and Ethnic Mobilization*. Routledge University Press.
- Phillips, A., & Reus-Smit, C. (2019). *Culture and Order in World Politics: Diversity and Its Discontents*. Cambridge University Press.
- Pollo, S., & Puto, A. (1981). *The History of Albania: From Its Origins to the Present Day*. Routledge CRC Press.
- Pop-Eleches, G., & Tucker, J. A. (2020). Communist Legacies and Left-Authoritarianism. *Comparative Political Studies*, 53, 1861-1889. DOI: 10.1177/0010414019879954
- Roeder, P. G. (1991). Soviet Federalism and Ethnic Mobilization. *World Politics*, 43, 196-232. DOI: 10.2307/2010471
- Rogers, A. R. (1988). Does Biology Constrain Culture. *American Anthropologist*, 90, 819-831. DOI: 10.1525/aa.1988.90.4.02a00030
- Schwander-Siever, S., & Fischer, B. (2002). *Albanian Identities: Myth and History*. Indiana University Press.
- Slezkine, Y. (1994). The USSR as a Communal Apartment, or How a Socialist State Promoted Ethnic Particularism. *Slavic Review*, 53, 414-452. DOI: 10.2307/2501300
- Stalin, J. V. (1913). *Marxism and the National Question*. Retrieved from <https://www.marxists.org/reference/archive/stalin/works/1913/03a.htm>
- Stickney, E. (1926). *Southern Albania Or Northern Epirus in European International Affairs, 1912-1923*. Stanford University Press.
- Stokes, L. (1987, 11). *Soccer Players Tell of Escape from Homeland*. Retrieved from <https://www.upi.com/Archives/1987/11/11/Soccer-players-tell-of-escape-from-homeland/7811563605200/>
- Suny, R. G. (1993). *The Revenge of the Past: Nationalism, Revolution, and the Collapse of the Soviet Union*. Stanford University Press.
- Thirty-three Albanians Flee by Boat to Greek Island of Corfu*. (1971, 8).
- Treisman, D. S. (1997, 1). Russia's "Ethnic Revival": The Separatist Activism of Regional Leaders in a Postcommunist Order. *World Politics*, 49, 212-249. DOI: 10.1353/wp.1997.0006
- Voigtländer, N., & Voth, H.-J. (2012, 8). Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany. *The Quarterly Journal of Economics*, 127, 1339-1392. DOI: 10.1093/qje/qjs019

A Albanian Civil Registry, 1930-45, Dataset Description

Link: Each record available on *Familysearch.org* includes a web link that redirects a user of the dataset to the webpage that hosts a scanned image of the record linked to the individual row. The link should always take a user to the first page where the household is located, but usually the row also contains information drawn from the succeeding page related to household changes after 1930. I learned during fieldwork that the paper used for these records were too wide for any existing scanner in the country,

until a makeshift one was made by a government bureaucrat in the 2000s specifically for preserving the records. The *Familysearch.org* team took two photos for each page.

Household ID: The household ID is an identifier the Albanian government introduced with this registry for the household of each village.

Name: This is the first name of the individual. Regarding spelling, I aimed to remain faithful to the spelling of the name recorded in the original document. For instance, if the first name was in Greek script, I recorded the name in Greek script.

Of note, however, to increase the utility of the dataset, in the cases in which the same person was spelled with slightly different spellings in relation to members of the same household, I assumed the census official(s) made an error and used the same spelling as initially recorded in this column.

Surname: This is the family name or the last name. Notably, in several cases, we observed different surnames within the same household. This is mostly related to people having moved from other families to be married within the household but retaining their last name. In retrospect, this is not surprising because in this period, large households made up of several smaller family units was the common practice (Kaser, 2014).

Father's name: This is the father's name, which was often used as a key identifying feature of individuals in the Ottoman period. The father's name, combined with an individual's name, continues to function as a key identifier in many Balkan countries' records akin to US-based social security numbers. In the cases missing a father's name, the individuals were termed "natural children" by the census officials.

Mother's name: This name is more often missing or unable to be transcribed due to damage to the document and/or perhaps because census officials prioritized the father's name, in line with previous Ottoman record-keeping practices.

Gender: This column denotes the sex of the individual (male or female).

HoH Relation: This variable denotes the relation of each individual to the head of the household (HoH). Each household has a household head - typically the most senior male. However, there are quite a few cases of female household heads, especially widowed women. This creates a complication for the cases in which the household head died after 1930 and then, was replaced by someone else, which also necessarily changes the relation of everyone else in the household. For purposes of my research, I retained the relations with the original household head of 1930, even after their deaths.

Birth Year: This is the year the individual was born. There is a noticeable pattern of age heaping observed in the dataset through which the census officials rounded their ages to a multiple of five, thus their year of birth usually ends with a zero or a five. This occurred especially among older people and women, indicating generally an innumerate society with no pre-existing records for births. The younger individuals have a considerably less pronounced heaping pattern.

Birthplace: As I did for first name spellings, my general practice was to retain the originally recorded spelling of the toponym. Future users of the dataset may wish to collapse spellings, however, for places which are clearly the same, i.e., Athena v. Athens.

Civil status: This column denotes whether or not the person is married at the point of the registration process. If the person had gotten married but remained in the same household, the public official responsible for this record would ideally make note of this. If the person got married and joined a new household, this was ideally recorded as such in the out-migration column. In practice, this data is most reliable at the initial point of registration (in 1930) and became less reliable over time (until 1945).

Literacy: This variable denotes whether the individual was literate in reading and writing in any language. In quite a few cases, no information was recorded in the literacy column at all. These villages seem to have high quality literacy data: Muzinë, Dhivër, Llupsat, Qesarat, Lukovë, Piqeras, Ninat, Markat, Janjar, Shalës, Leshnicë e Siperme. The rest of the villages seem to suffer from inconsistencies or missing data in this column.

Occupation: This denotes the occupation of the individual. Similar to the literacy column, the quality of the recording is variable from one village to another. A high percentage of individuals have their occupations recorded as "workers," but it is unclear what type of work this entailed. For women, many of them were recorded as housewives. I did not note this down, since by traditional standards, all women in the household of working capacity had household duties. I only noted a woman's occupation if it was different from that of a housewife. For children, many are recorded as pupils, meaning that they were attending school, usually at the primary level.

Religion: This column records the religious identity of each individual. Most villages have the same religious identity for all families and individuals, with some important and interesting exceptions. The village of Markat has a mostly Muslim identity, but also has a minority Orthodox Vlach population. This population were largely nomadic shepherds and completely ignored by Hoxha's MRE perhaps due

to their nomadic character. There are religiously mixed families as well, where the wife was Orthodox Christian while the husband was a Muslim. In each of the mixed cases, regardless of whether the HoH was Christian or Muslim, all of the children were recorded as Orthodox Christian. This transmission trend toward Orthodox identity indicates a generally higher perceived status of Orthodoxy in this region. There were other cases of mixed religions within the same household, but mostly in the context of servants working for someone of a differing religion.

Death Date: This column is filled with a year if the individual died after 1930 and before 1945. The vast majority of the deaths are recorded for years prior to 1944, but there are some cases of later deaths as well indicating local record-keeping practices of updating this pre-communist record varied.

Village In: This column denotes the village from which the person migrated to the listed location.

Village Out: This column relates the villages the individual migrated to from the listed location. In cases in which the Year Out Migration column is filled but the Commune Out column is empty, a data set user can assume this means that the individual came from another household of the same locality.

Year In Migration: This is the year in which the individual migrated to another household. This is a decent measure for status change because the registry officials recorded information in this category even when individuals married someone in the same household. However, this can create confusion if a user of the dataset wants to look only at actual migration. To distinguish this, a dataset user can examine the Year Out Migration data column to determine whether there is a male and female who have the same year of migration.

Year Out Migration: This is the year in which the individual migrated from the recorded household to another. The individual might have moved to another household of the same locality or a different locality within the country or abroad. In cases the Year In Migration column is filled but the Village In column is empty, a user should assume it means that the individual came from another household in the same village. Future users of the data set could create a new column, denoting within-village marriages with these individuals.

Notes: I introduced this category in the data. In cases of a notable figure mentioned in another context, we included links to the additional relevant information.

B Villages Names Over Time

C Albanian Cemetery Dataset Description

Image Identifier This is the unique web link to the photo of the grave from which the information is reproduced in individual rows of the data set.

Village Name Here we record the name of the village in which the site was recorded. We use the Albanian language spelling first and follow that by other spellings of the same name to allow for maximum overlaps and navigability of the data.

District Here we record the district as recognized by Albania’s 2022 government lists.

Geolocation Here we record the coordinates attached to the photo.

Site type Each picture should be of a “site” or memorial associated with an individual cemetery. This column denotes what type of site is associated with the record (cemetery/ public square/church/mosque/political party office).

Script of written info Here we recorded which script is displayed on the grave (Greek/Latin/Cyrillic/none). Latin denotes the Albanian language and Cyrillic denotes Macedonian.

Gender identity Here we record whether the person represents the male or female identity. This can be inferred from the first name, the photo of the person and/or the spelling of the last name.

Last name This is the last name or surname of the memorialized individual. We record the name in the script displayed, whether Greek, Latin, or Cyrillic.

First name This is the first name of the memorialized individual. We recorded the name in the script displayed on the memorial (whether a grave or a monument). We used the name of saint for churches or figure associated with mosque.

Birth year This is year the represented individual was born. For sites/monuments, we put here the year the memorialized figure was born and if that’s not available, the year the site was constructed.

Orthodox symbol Here we record if there is an Orthodox symbol, such as a cross, as part of the grave.

Muslim symbol This column indicates whether a Muslim symbol, such as a hilaal (moon and crescent), is part of the grave.

Communist symbol This column indicates whether there is communist symbol, such as a red star, as part of the grave.

Double eagle This column indicates whether there is a double eagle displayed as part of the grave.

Other written text If the grave displays a eulogy related to identity, we transcribe it here.

Translation of eulogy Here we also translate eulogies relevant to our study from Greek and/or Albanian to English.

Recent visit? This variable indicates if someone has left an identity marker while recently visiting the grave, such as an additional cross, hilaal, or flowers. This column can be used later to parse identity transmission practices.

How complete is the written info? This is where the RA records whether the recorded information is complete. (complete / incomplete)

How is it incomplete? This column describes which part(s) of the data are missing or unable to be deciphered.

Any other info? This column denotes anything additional the RA thinks the PI might want to know, especially related to political dynamics and/or social identities.

RA(s): This column lists the RA(s)' names who recorded the row of data.

Reviewer notes: Reviewer notes. This column is for raising any confusion a reviewer of the dataset might have with a specific row of data.

PI notes: Miscellaneous notes are kept here, especially regarding decisions made about confusing records, for future reference.