The Geopolitics of Repressions

Martin Kosík

April 4, 2019

Abstract

I study how geopolitical concerns influence attitudes of a state toward its minorities. I exploit the Hitler's rise to power in 1933 as an exogenous shock to Soviet-German relations. Using the digitized archival data on 2.7 million individual arrests by the Soviet secret police, I apply difference-in-differences and synthetic control method to estimate how changing geopolitical relations influenced repressions of Germans in the USSR. The estimates of both models imply that that the arrests of Germans relative to other minorities increased by approximately 2% after 1933.

Contents

Introduction

What determines the attitude of a state toward ethnic minorities within its borders? Why are some minorities accommodated or assimilated and others are politically excluded and repressed? Furthermore, why does the position of a state toward its minorities change in time? For example Soviet Union largely accommodated its minorities by in 1920s but it heavily repressed them in the campaigns of mass terror 10 years later.

Mylonas (2013) argues that the geopolitical concerns play an important role. Specifically, a state is likely to choose repression and exclusion if the ethnic minority's country of origin is seen as an geopolitical enemy. The minority is then viewed by the state as unreliable and as a potential fifth column of the foreign country.

We test this hypothesis on the case of German minority in Soviet union. In 1933, Hitlers rise to power changed Germany from a neutral actor to ideological and geopolitical enemy in the perspective of the Soviet Union. We can then see how the repression changed before and after 1933 and compare it with other minorities. In particular, we use the individual arrests by the Soviet secret police as a dependent variable and employ the difference in difference strategy and the synthetic control method.

1 Literature Review

Existing literature on repressions has focused mostly on their consequences and legacies (Rozenas et al., 2017; Lupu and Peisakhin, 2017; Zhukov and Talibova, 2018). As far as the strategic use of repressions by the state is studied, it is usually in relation to domestic factors such as institutions and economic shocks (Davenport, 2007; Greitens, 2016; Blaydes, 2018) with less attention being given to external forces.

Davenport (2007) finds that democracy is correlated with lower levels of repression. However, it is mainly free electoral competition rather than constraints on power of the executive that accounts for this negative effect. Greitens (2016) links the severity of repression to the threat from dictator's inner circle. A dictator who fears that he would be deposed in a coup rather than in popular uprising will fragment their coercive apparatus in order to weaken the power of a potential challenger from within. The weakened secret police will be, according to Greitens, more likely to use violence since it fails to identify the transgressors and cannot effectively deter dissent. Other scholars see repression as a substitute for co-option (Wintrobe, 1998; Svolik, 2012). Instead relying on the threat of persecution, an authoritarian ruler might buy the loyalty of the population by distributing rents to the supporters of the regime usually through the party apparatus. Negative economic shocks can then increase repression since the rents are no longer available. Blaydes (2018) illustrates this on the case of Iraq under Hussein where lower oil prices

Furthermore, Blaydes (2018) presents a theory attempting to explain different levels of repression across ethnic minorities within a country. She argues that the nature of repression depends on the legibility of the ethnic group to the state.[1] Since the state coercive institutions cannot reliably identify transgressors in less legible population (because of, for example, greater cultural and linguistic distance), they will more tend to resort to collective punishment. The logic behind this is that the members of the group will police its members to avoid collective punishment.

Our research also contributes to the literature studying factors that influence position of a state towards its ethnic minorities and under what conditions conflict is likely to occur. Size and and distribution of ethnic groups have been emphasized. Several scholars pointed out that states with large number of ethnic groups are more likely to violently repress calls for autonomy or secession to discourage other ethnic minorities from making similar demands in the future (Evera, 1994; Toft, 2005; Walter, 2009). Furthermore, Toft (2005) argues that geographically concentrated groups tend to view their ethnic homeland as indivisible and non-negotiable issue which increases the likelihood of violent conflict. However, these approaches fail to explain changes in state's attitudes to minorities over

[1] The term legibility in this context means ability of a state to identify individuals in a given population and gather information on them.

short periods of time when the size and distribution of ethnic groups remains roughly constant.

More recently, the role of international factors have received greater attention. Butt (2017) argues that response of a state to a secessionist movement depends on external security environment and outside actors. Specifically, a state located in a war-prone region is more likely to suppress demands for secession because loss of territory and population would make it vulnerable to a potential future attack. Furthermore, a state responds with more violence if a separatist movement receives a support from an external power since the outside assistance makes the secessionists stronger. In addition to these strategic reasons, Butt emphasizes that receiving external support incites a strong feeling of betrayal to the central state.

As was mentioned, Mylonas (2013) puts forward a theory explaining how geopolitical relations influence the attitude of a state towards its minorities. The model features an ethnic minority living within a host state and an external power. Moreover, Mylonas distinguishes between hosts states with revisionist foreign policy which want change the international status quo (e.g. because they gained power or lost territory in the past) and host states that prefer the current international order.

The predictions of the theory are summarized in the table 1. First, if a minority group is not supported by any external power, the theory predicts that the host state will pursue policy of assimilation towards the group to "immunize" it from possible future agitation of external powers. Second, if an ethnic minority is supported by geopolitical ally then accommodation is likely since more repressive policies towards the minority could jeopardize the alliance. Third, theory predicts assimilation if a minority is supported by an geopolitical enemy and a host state pursues non-revisionist foreign policy because exclusionary policies could trigger new hostilities threatening the status quo. Finally, support of an external enemy combined with revisionist foreign policy will likely lead to exclusion of a given ethnic group since it is view as a potential "fifth column" of the external power.

We can apply the theory to our case. First, the Soviet Union was arguably a revisionist state. The Bolsheviks had to accept large losses of territory under the Treaty of Brest-Litovsk in 1918 so that they could focus on fighting in Russian Civil War. Thus, the USSR would certainly prefer to change the international status quo. Second, the German-Soviet relations were neutral of even moderately friendly prior to 1933 but turned hostile after the rise of Hitler (described in greater detail in subsection 2.1). Therefore the theory predicts that the policy of the Soviet state towards the German minority should change from accommodation to exclusion. Mylonas (2013, p. 22) defines exclusion as "policies that aim at the physical removal of a non-core group from the host state." The Soviet political

		External Power Support		
		Yes		No
		Interstate Relations		Assimilation
		Ally	**Enemy**	
Host's State	**Revisionist**	Accommodation	Exclusion	
Foreign Policy	**Status Quo**	Accommodation	Assimilation	

Table 1: Theoretical predictions of Mylonas (2013)

repression which usually featured either outright execution or a term in a labor camp in the Far East of the country fits this description well. Thus, the theoretical expectation is that repressions of Soviet Germans relative to other minorities should increase after 1933.

Our main contribution is empirical test of this theory using a credible identification strategy. Most studies presented in this review test their hypotheses only by qualitative comparison of selected cases. Quantitative research usually involve only cross-sectional regressions based on categorical dependent variables.

For example, Mylonas (2013) tests his theory with data on the post-World War I Balkans where the nation-building policies (categorized into 3 groups: accommodation, assimilation and exclusion) toward 90 ethnic groups are a dependent variable and information on their support by external powers is an explanatory variables (together with other control variables). However, the results of the cross-sectional regression, used in the study, might easily be biased due to omitted variables or reverse causality and we believe that our approach offers cleaner identification.

McNamee and Zhang (2019) is methodologically and thematically closet study to ours. They analyze how the 1958 split in Soviet-China relations affected the demographic composition of the population in the Soviet-Chinese border regions. Using difference-indifference strategy, they find that, after the split, both states supported expulsions of the minority group and sponsored immigration of the majority group but only in border regions without significant natural boundary (e.g. high mountains). They conclude that the states use demographic engineering as a way to protect their vulnerable border against a hostile power.

2 Historical Background

In this section, we provide brief historical context for selected topics. Specifically, we first describe changing geopolitical relations of Germany and the Soviet from 1920 to 1941. In the next subsection, we provide brief overview of the most important aspects of Soviet political repression. Finally, evolution of the Soviet policy towards its ethnic minorities is summarized.

2.1 German–Soviet Relations in the Interwar Period

The relations between Weimar Germany and Soviet Union can be characterized as neutral or even cooperative. Both countries were somewhat isolated in the international system dominated by western powers (Great Britain, France, USA) and sought to find allies. The good relations were first established by the Treaty of Rappalo in 1922 in which both countries renounced the territorial and financial claims against the other and agreed to secret military cooperation (Gatzke, 1958) and then reaffirmed by the Treaty of Berlin in 1926. Furthermore, a trade treaty was signed between the two countries in 1925 (Morgan, 1963).

Hitler was named chancellor on 30 January 1933 and effectively become a dictator on 24 March 1933 by the passing of the Enabling Act. The relations with Soviet Union quickly turned hostile for several reasons. First, Hitler called in *Main Kampf* for Germany to obtain *Lebensraum* (living space) in the east, presumably at the expense of the Soviet Union and he often spoke of Judeo-Bolsheviks. Moreover, Hitler soon after his rise to power banned the German Communist Party and started to persecute its members (Haslam, 1984).

The opposition to fascism led to change in policy of the Communist International (Comintern) with appointment Georgi Dimitrov as its general secretary in 1934. The Communist parties in democratic countries were now encouraged to form coalitions (Popular Fronts) with social democratic parties to prevent rise of fascism, in contrast to the previous aggressive and uncompromising approach. This policy was affirmed by the Seventh World Congress of the Comintern in 1935 (Haslam, 1979).

The newly formed Popular Front coalitions won elections and entered government in some European countries including France and Spain. In Spain however, the coup of nationalists against the new government in 1936 sparked a civil war. The Soviet Union heavily supported the republican government, while Germany supplied the nationalists which further increased the tensions between the two countries.

As a response, Japan and Germany signed the Anti-Comintern Pact in 1936 in which they committed to co-operate for defense against communistic disintegration. Meanwhile in the Soviet Union, many people were persecuted for alleged cooperation with Germany including leading general Mikhail Tukhachevsky.

The orientation of German foreign policy began to shift in spring of 1939. Until that point, Hitler tried to court Poland to join the Anti-Comintern Pact against the Soviet Union (Weinberg, 2010, chapter 26). But Poland repeatedly refused and the German army began planing for the invasion of Poland in April 1939 (Kotkin, 2017, p. 621). However, France and Great Britain granted security guarantees to Poland in March 1939. Hitler thus tried to negotiate neutrality of the Soviet Union in war to avoid simultaneously facing Western powers, Poland and the Soviet Union. Soviet neutrality was potentially beneficial for Stalin too. A long and costly war would weaken the both the capitalist and fascist enemies of the Soviet Union. Moreover, Stalin believed that conditions of war could bring about socialist revolutions in those countries just as in Russia in 1917. After brief negotiations, on 23 August 1939 the Molotov-Ribbentrop pact was signed between Germany and the USSR which guaranteed non-belligerence between the two countries. In addition, a secret protocol of the treaty marked the German and Soviet spheres of influence in Eastern Europe.

The pact of the two former ideological enemies caused great shock and astonishment both among Party officials and ordinary people. Victor Kravchenko (1947, p. 332), a Soviet official who later defected to the US, described in his memoir the disbelief upon hearing about the pact

> There must be some mistake, I thought, and everyone around me seemed equally incredulous. After all, hatred of Nazism had been drummed into our minds year after year. The big treason trials [...] have rested on assumption that Nazi Germany and its Axis friends [...] were preparing to attack us.

Another party official later recalled that “it left us all stunned, bewildered, and groggy with disbelief” (Robinson and Slevin, 1988, p. 137).

Nazi Germany attacked Poland on 1 September 1938 from the west and shortly after that, on 17 September, the Red Army invaded the eastern part of the country. As was agreed in the pact, Poland was partitioned between Germany and the Soviet Union. However, the mistrust between the two countries was still present as evidenced by a violent clash of German and Soviet troops near Lwów on 20 September.

Hitler enjoyed major success in the first years of the war. By summer 1940, German forces defeated French army and annexed Denmark and Norway. However, German industry was severely lacking raw materials needed in war effort against Britain which, according to some historians, motivated Hitler to invade resource-rich Soviet Union (Tooze, 2008). The German attack on the Soviet Union on 22 June 1941 ended 2 years of fragile cooperation.

2.2 Soviet Political Repressions

The Soviet Union had massive coercive apparatus. The Soviet secret police (which was throughout the years named the Cheka, OGPU, NKVD, MVD and the KGB)[2] employed at its height (1937–1938) 270,730 persons (Gregory, 2009, p. 2). The political repressions were usually carried under Article 58 of the Criminal Code. The article 58 punished counter-revolutionary activities which included treason, espionage, counterrevolutionary propaganda or agitation and failure to report any of these crimes. Most In practice, this broad definition meant that anyone regarded as politically inconvenient could be arrested and prosecuted.

During the mass operations, the central office of the NKVD would typically set quotas for the number of arrests which the regional branches were supposed to reach and exceed (Gregory, 2009, chapter 6). The local NKVD officer had to decide themselves who to target to meet the quotas.

The sentences were in most cases issued extrajudicially by so-called “troikas”, three-person committees composed of a regional NKVD chief, a regional party leader, and a regional prosecutor. The NKVD chief usually dominated the process as party leaders sometimes feared that they themselves would be targeted (Snyder, 2011, p. 82). Only rarely was a person acquitted from his charge. The most common sentences for political crimes in the Stalinist period were execution and prison term in a labor camp (Gulag) (Gregory, 2009, p. 21). A term in the Gulag of less then 5 years was considered lighter sentence in these cases.

With the rise in repressions in the 1930s, the Gulag system significantly expanded. At its height, it consisted of at least 476 distinct camp complexes each containing hundreds of prisoners. The Gulag system offered the Soviet state cheap source of labor that produced substantial amount the country’s coal, timber, and gold supply. The mortality of prisoners was high due to heavy work, malnutrition and cold climate (**?**).

The death of Stalin in 1953 marked a start of decline in political repressions in the USSR. The new Soviet leader, Nikita Khrushchev, denounced Stalin and the mass repressions of his period in his speech *On the Cult of Personality and Its Consequences* in 1956. The suppression of dissent continued in the Khrushchev and Brezhnev era but in much milder form. Khrushchev gradually dismantled the Gulag system, granted amnesty to many political prisoners and started the process of rehabilitation of victims of the Stalinist period although they were limited to only some categories of victims and offences

[2] We will refer to the Soviet secret police as the NKVD in this text since this was the name of the agency for the largest part of the period of our interest

(?Dobson, 2009).

2.3 Ethnic Minorities in the Soviet Union

The Soviet Union was from its inception a multi-ethnic state. According to the 1926 Census, the Russians made up only half of the total population.[3] Among other large ethnic group were Ukrainians, Belorussians and Kazakhs. A significant fraction of citizens of the USSR belonged to ethnic groups with their own independent states including Polish, German, Estonian, Latvian, Lithuanian, Finish, and Greek minorities. The Bolshevik elites were aware of the multi-ethnic nature of their newly formed state and wanted to avoid a perception of the Soviet Union as a project of Russian imperialism. Furthermore, the Bolsheviks hoped that they could exert political influence in countries with cross-border ethnic ties to Soviet diaspora nationalities by promoting the interests of minorities in the USSR (this was know as the "Piedmont Principle").

As a consequence, the Soviet policy towards its ethnic minorities in the 1920s was largely accommodating (Martin, 2001). The languages and culture of minorities were promoted and minorities were encouraged to enter local governments and party structures (so-called *korenizatsiya* policy). Some minority groups were well represented even in the NKVD (Gregory, 2009, p. 25). In some cases Autonomous Soviet Socialist Republics (ASSR) were established (including Volga German ASSR) which had given the regional minorities certain degree of independence.

This attitude changed drastically in the 1930s. First, the *korenizatsiya* policy started to be reversed in the 1932. From 1934, the NKVD started to deport ethnic minorities from the state frontier zone in Eastern Europe. This involved forced resettlement of 30 000 of Ingermanland Finns and tens of thousands of Poles and Germans to Kazakhstan and West Siberia (Polian, 2003, p. 95). In 1937 and 1938, the NKVD conducted mass operations specifically targeted at minorities with cross-border ethnic ties. Poles, Latvians, Germans, Estonians, Finns, Greeks, Chinese, and Romanians were arrested in large numbers as supposed spies and saboteurs of foreign governments. More than 320 000 people were arrested in the national operation out of which about 250 000 were executed (Martin, 1998, p. 855). The Polish operation claimed the highest number of victims (110 000 deaths) but many .

The persecutions further escalated with the World War II. Following the German invasion into the Soviet Union in 1941, Stalin ordered deportation of about 430 000 Soviet Germans (most of them living in Volga German ASSR) into Kazakhstan and Siberia (Polian, 2003, p. 134). Similar "preventive" deportation followed for Finns and Greeks

[3] The census data were obtained Institute of Demography of the National Research University Higher School of Economics

as well. Between 1943-1944, forced resettlement of another six ethnic groups (Karachais, Kalmyks, Chechens, Ingushetians, Balkars, and Crimean Tatars) were carried out for alleged or actual cooperation of some of these minorities with the German troops (even if many more served in the Red Army).

Table 2: Missing Data by Variable

Variable	Number of Missing Obs.	Percent of Missing Obs.
Ethnicity	1 507 177	55.73
Date of Arrest	1 650 912	61.04
Date of Process	943 108	34.87
First Name	14 006	0.52

3 Data

Our data on Soviet repressions come from the Victims of Political Terror in the USSR database by the Russian human rights organization Memorial (2017).[4] The Memorial data has already been used in empirical research by Zhukov and Talibova (2018) to estimate the effects of the Soviet repressions on the current political participation. The main sources of the Memorial lists are declassified Russian Interior Ministry documents, prosecutor's offices and the Commission for the Rehabilitation of Victims of Political Repression, and "Books of Memory". Vast majority records in the database are individual arrests under Article 58. This means that the victims of other repressive activities of the Soviet state such mass forced migration, counter-insurgency operations of the Red Army during Russian Civil War or famines are mostly excluded. Therefore our research focuses on one particular type of repression (individual arrests by the NKVD).

Nevertheless, even if we restrict ourselves to the individual arrests, the Memorial database is still not complete. At the time of our access to the database, the databes contained 2.7 million records which is approximately 70% of 3.8 million people convicted under Article 58 between 1921 and 1953.

Missing data presents another major challenge. Table 2 shows how many values are missing for the variables of our interest. We can see that information on ethnicity is not available for more than half of all observations. In contrast, a surname is recorded for every arrest in the dataset and first name is missing only for negligible fraction of observations. The availability of information on names enables us to use them to infer missing ethnicity of an individual. In particular, we will train a Naive Bayes classifier on the 1 197 373 with known ethnicity and use the model's predictions to impute the ethnicity for the remaining 1 507 177 observations (the details are described in the subsection 4.1).

Date of arrest, which is also necessary for our analysis, has even higher rate of missingness than ethnicity.[5] One solution, albeit only partial, might be to use the date of

[4] The database can be accessed and searched from http://base.memo.ru/ (new version) or at http://lists.memo.ru/ (older version). However, we downloaded the file with every record from the database from https://github.com/MemorialInternational/memorial_data_FULL_DB/blob/master/data/lists.memo.ru-disk/lists.memo.ru-disk.zip

[5] We consider a date missing only if not even year of arrest is available. Many observations have

Table 3: Missing Dates of Arrest and Process

	Date of Arrest	
Date of Process	Missing	Present
Missing	747 419	195 689
Present	903 493	857 949

Table 4: Missingness of Ethnicity and Dates of Arrest and Process

	Date of Arrest or Process	
Ethnicity	Missing	Present
Missing	594 975	912 202
Present	152 444	1 044 929

process (where it is available) to extrapolate the missing date of arrest. As is shown in the table 3, we could impute this way the missing date of arrests for 903 493 observations for which we have the date of process. Total number of arrests used for our analysis would therefore increase from 1 053 638 to 1 957 131. However, there remains 747 419 observations for which neither the date of arrest nor the date of process are known.

After imputations of ethnicity and the date of arrest had been applied, we created our main dataset by counting number of arrest for each ethnicity by year.

With 38 minorities and 40 time periods (from 1921 to 1960) we have 1520 observations in total. Total number of arrests for each ethnicity is provided in the table 6 and the plot of arrest by ethnicity and year (after applying the transformation $\log(1 + y_{it})$) is shown in figure 7, both in the appendix.

In addition to data on repressions, we also obtained some information on some characteristics of the ethnic groups in our dataset. Specifically, we acquired total population of the ethnic groups and their urbanization rate from 1926 Soviet Census from the Demoscope website.[6] For each ethnic group, we also calculated the cladisitc similarity of its language to Russian based from Glottolog language trees (Hammarström et al., 2018). Cladistic measure of linguistic similarity counts the number of shared branching points between the two nodes on a language tree and it has been used by Fearon (2003) and Dickens (2018) among others. The full data are provided in the table 11 in the appendix.

information on year of arrest even though the exact month or day are missing. We do not categorize these observation as having missing date of arrest since our fundamental unit of analysis is a year and thus month and day are not of high interest to us. The same applies to the date of process.

[6] It is available online at http://www.demoscope.ru/weekly/ssp/ussr_nac_26.php

4 Imputation of Missing Data

4.1 Inferring Ethnicity from Names

In this section, we explain our method for predicting ethnicity of an individual from his or her names. Using names for imputing ethnicity has several advantages. First, full name is available for every individuals in the dataset. Second, names have been shown to be highly predictive of ethnicity in a variety of applications (Mateos, 2007; Hofstra et al., 2017; Hofstra and de Schipper, 2018).

Given the high number of predictors, we need a model that is not computationally demanding but at the same time achieves reasonable level of prediction accuracy. Naive Bayes classifier meets these criteria and has been for this reason used in wide range of applications including text classification (Gentzkow et al., 2019).

4.1.1 Naive Bayes Classifier

Let $\boldsymbol{x} = (x_1, x_2, x_3)$ be features used for predicting ethnicity, that is a person's first, last, and patronymic (given after father's first name) names. Using Bayes theorem, we can express the probability that particular observation belongs to ethnic group E_k given its features as

$$p(E_k \mid \mathbf{x}) = \frac{p(E_k)\ p(\mathbf{x} \mid E_k)}{p(\mathbf{x})}, \tag{1}$$

in other words, the posterior probability is proportional to the product of prior probability and likelihood. Assuming conditional independence of features allows us to substitute $p(\mathbf{x} \mid E_k)$ such that we get

$$p(E_k \mid \mathbf{x}) = \frac{p(E_k)\ \prod_{i=1}^{3} p(x_i \mid E_k)}{p(\mathbf{x})}. \tag{2}$$

All terms in this equation now can be estimated from the data: the prior probability $p(E_k)$ as a proportion of E_k in the data, $p(x_i \mid E_k)$ as a proportion of people with name x_i in the ethnic group E_k and $p(\mathbf{x})$ simply calculated such that the sum of $p(E_k \mid \mathbf{x})$ for all k is one. The Naive Bayes classifier then chooses the ethnicity with the highest posterior probability as its prediction, that is

$$\hat{y} = \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}}\ p(E_k) \prod_{i=1}^{3} p(x_i \mid E_k). \tag{3}$$

However, one potential issue is that whenever a likelihood of a certain feature is estimated to be 0 then the posterior probability is always 0 regardless of the prior or the likelihoods of other features. For example, suppose that a person has a typical German first name but a rare surname which does not appear in the training set at all. Then the useful information contained in the first name will be completely ignored since the 0 likelihood of the surname will override any other value and we will end up with the posterior probability of zero for all ethnic groups.

To address this problem, we apply Laplace smoothing. For every ethnicity k, let c_j be number of people with a name j and N be total number of people. Without applying any smoothing, we would estimate the likelihood $p(x_i \mid E_k)$ simply as a relative frequency, i.e. $\hat{\theta}_j = \frac{c_j}{N}$. With Laplace smoothing, we estimate the likelihood $\hat{\theta}_j$ as

$$\hat{\theta}_j = \frac{c_j + \alpha}{N + \alpha d} \qquad j = 1, \ldots, d \tag{4}$$

where parameter $\alpha > 0$ is a smoothing parameter. This ensures that for any finite value of N, $\hat{\theta}_j$ will never be zero.

It is important to note that the conditional independence assumption often does not hold in the data and the estimated posterior probabilities therefore have to be taken with a grain of salt. However, our main goal is the best out-of-sample accuracy of the model's predictions. In this respect, Naive Bayes classifier have been shown to perform well in many applications, despite its often violated assumptions (Domingos and Pazzani, 1997).

4.1.2 Adjusting for Unbalanced Prediction Accuracy

To reliably asses the out-of-sample performance of our model, we used 10-fold cross-validation on the training dataset. That is, the data are first randomly split into 10 groups, Then the model is fit to the data from to the data .. Using this method, we get the overall 75.4% accuracy. Nevertheless, it varies significantly by ethnicity. The specificity and sensitivity by ethnic group are provided in the table 7. Some ethnic groups with distinctive names such as Chinese, Korean and German are classified fairly well with sensitivity higher than 75%.

Thus, there are biases in our predictions. However, we can try to adjust for them. Let P_{it} be the number of people with predicted ethnicity i arrested at time t, R_{it} be actual the number of people with ethnicity i arrested at time t, α_i and β_i be sensitivity and specificity of our classifier for ethnic group i and N_t be the total number of arrests at time t. Then the predicted arrests of a given ethnicity are sum of true positives and false positives, that is

$$P_{it} = \alpha_i R_{it} + (N_t - R_{it}) \cdot (1 - \beta_i) \tag{5}$$

We are interested in R_{it} but we only directly observe P_{it} and N_t. However using simple algebra, R_{it} can be expressed as

$$R_{it} = \frac{P_{it} - N_t\,(1 - \beta_i)}{\alpha_i + \beta_i - 1} \tag{6}$$

We will refer to this method of correcting predictions as parsimonious adjustment. The parameters α_i and β_i are not known to us but we can use their estimates from the cross-validation on the training data. This assumes that the these parameters do not differ significantly for the training and test data. But this might not be the case. Suppose, for example, that Armenians are often misclassified as Chechens and that the number of Armenians in the data with missing ethnicity is disproportionately higher than in the data with information on ethnicity. Then the cross-validated specificity for Chechens in the training set will underestimate the specificity in the test set because it does not take into account higher proportion of Armenians.

Fortunately, we can address this potential bias by more building complex modeling. First for all ethnic groups i and j, we define the misclassification rate b_{ij} as share of people with ethnicity j that are classified as i. Notice that for $i = j$, the misclassification rate is simply prediction accuracy for ethnicity i. It follows from the definition of the terms that predicted number of arrests for ethnic group i at time t, P_{it}, is equal to

$$P_{it} = \sum_{j=1}^{K} b_{ij} R_{jt} \qquad i = 1, \ldots, K \tag{7}$$

This equation can be expressed in matrix form as

$$\mathbf{P}_t = \mathbf{B} \cdot \mathbf{R}_t \tag{8}$$

where $\mathbf{P}_t = (P_{1t}, \cdots, P_{Kt})$, $\mathbf{R}_t = (R_{1t}, \cdots, R_{Kt})$, and $\mathbf{B} = (b_{ij})_{i=1,\ldots,K,\, j=1,\ldots,K}$. To express $\mathbf{R}_t$, we just apply basic linear algebra

$$\mathbf{R}_t = \mathbf{B}^{-1} \cdot \mathbf{P}_t \tag{9}$$

We will call this method the full matrix adjustment. Compared to the parsimonious adjustment (in equation 6), this correction no longer assumes that the test set sensitivity and specificity be accurately estimated from the training set. The full matrix adjustment makes only somewhat weaker assumption that the train and test set misclassification rates are not significantly different.

4.2 Imputing Missing Date of Arrest

Our strategy for imputing the missing arrest dates is to predict it from the date of process. For this reason, we model the number of days between arrest and process and fit it to a subset of the data for which both dates are known. It is reasonable to expect that the average number of day from arrest to process could vary considerably throughout the years. Hence we use the year of process as a predictor to our model.

Nevertheless, we begin by examining the data with both dates available. The histogram for number of days between arrest and process (on scale of $\log_{10}(1+x)$) is shown in the figure 1. First, we can see that there is fairly large variance in the variable with number of days ranging from 0 to more than 1000.[7] Second, the transformed data seems to be following the normal distribution except for the density at 0 which is much higher than the normal model would predict. Moreover, the zero values are making the estimated mean of the normal distribution lower than would be appropriate for the positive values resulting in poor fit.

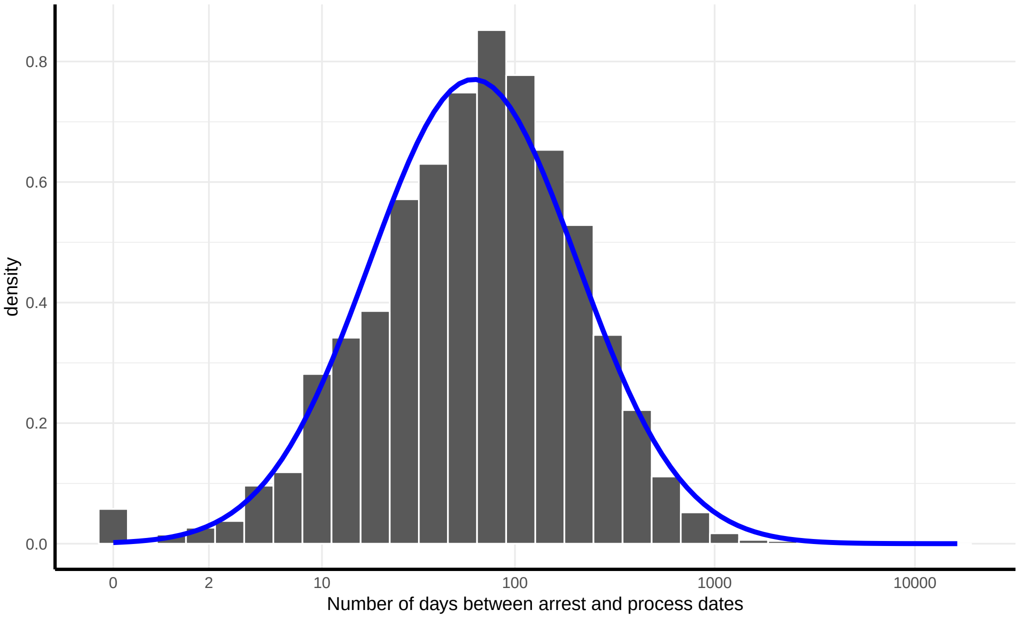


Figure 1: Histogram for Number of Days between Arrest and Process (shown on $\log(1+x)$ scale)

To address this problem, we model the zero and positive values separately in a two-stage process using a method described in Gelman and Hill (2006, p. 537-538). Let y be the number of days between arrest of an individual and his or her process and X be set of dummy variables indicating the year of process. We also define I^y as an indicator variable that equals 1 if $y > 0$ and 0 otherwise and y^{pos} to be only positive values of y

[7] Zero, of course, corresponds to both arrest and process being in the same day.

(i.e. $y^{\text{pos}} = y$ if $y > 0$). In the first stage, we predict I^y using logistic regression

$$\Pr\left(I_i^y = 1\right) = \text{logit}^{-1}\left(X_i \alpha\right). \tag{10}$$

In the second stage, simple log-linear regression is applied to predict only the positive values $y^{\text{pos}} = y$

$$\log\left(y_i^{\text{pos}}\right) \sim \text{N}\left(X_i \beta, \sigma\right). \tag{11}$$

We then fit the first model to the data where the exact dates of both arrest and process are available and the second model to the subset of the same data for which $y > 0$. The results of both of these models provided in the table 10 in the appendix. The years of process appear to be important predictors both in the first stage and even more in the second stage. However, the unexplained variance is still high making up about 76% of the total variability in the dependent variable in the second model.

We proceed to apply the fitted models to the missing data to get the predicted probability of y being positive and the mean value of y if it is positive. For each observation with missing date of arrest X_i, we then randomly draw from the Bernoulli distribution with $\text{logit}^{-1}\left(X_i \hat{\alpha}\right)$ as its parameter to obtain $\hat{I}_i^y$. We also draw from the normal distribution with mean $X_i \hat{\beta}$ and exponentiate the result to get $\hat{y}_i^{\text{pos}}$. Finally, the predicted number of days is calculated simply as $\hat{y}_i = \hat{I}_i^y \cdot \hat{y}_i^{\text{pos}}$.

The histogram of the imputed values is provided in the figure 8 in the appendix. The resulting distribution highly resembles the distribution in the figure 1 including the fraction of zero values indicating our model captures the actual data fairly well.

Another issue is that for significant number of observations we do not have the exact date of process but only year. In particular, while the year of process is recorded for all 903 455 observations where the date of arrest is to be imputed, the month of process is missing for 369 393 of them and the day for 390 174. To fill in the missing month, we take a random sample from all months with probability equal to the relative frequency of the months of process in the non-missing data between the years 1921 to 1960. Even simpler method is used to impute the missing days where we just randomly choose a day within given month with uniform probability. The imputed months and days of process are therefore only weakly informed guesses, nevertheless they enable us to proceed in the analysis.

The final step is to calculate the imputed date of arrest by subtracting the predicted number of days from the date of process (i.e. we go back in time by given number of days). Since we conduct the analysis with annual observations, we ignore predicted month and day of arrests keep only information on year. The number of arrests for each ethnicity by year (including the imputed years) is then counted for the period from 1921 to 1960 which forms our final dataset.

The resulting time series of all arrests with imputed years is plotted in the figure 9 in the appendix. Arrests with imputed dates seem to follow similar trends with the labeled data although there is slight divergence at the beginning and end of the series and in the 1930s.

5 Methodology

5.1 Difference-in-differences

Our main specification is the dynamic difference-in-differences model:

$$\log\left(1+y_{it}\right)=\sum_{k=1930}^{1939}\beta_k\,\mathrm{German}_i\cdot\mathrm{Year}_t^k+\lambda_t+a_i+E_i\cdot t\;+E_i\cdot t^2\;+u_{it} \tag{12}$$

where y_{it} is number of arrests of people with ethnicity i in year t, λ is year fixed effect, a is ethnicity fixed effect (both captured by respective dummy variables) and Year_t^k are dummy variables that equals 1 if its year k equals to t and 0 otherwise (except for 1939 which is equal to 1 if $t\geq 1939$ and zero otherwise) . The coefficients of interest are β_k. Prior to 1933 they capture the lead (anticipatory) effects used to test if pre-treatment trends are parallel. After 1933 they capture the dynamic lagged effects. The $E_i\cdot t$ and $E_i\cdot t^2$ term capture the ethnicity specific quadratic time trends. The inclusion of this term should not significantly change the coefficients, unless the results are driven by spurious correlation (Angrist and Pischke 2009).

We apply logarithmic transformation on y_{it} since it better fits the data (more in the results below). We use $\log\left(1+y_{it}\right)$ because some observations (although not many) have $y=0$. As discussed in Wooldridge (2015, p. 193), the percentage change interpretation is usually closely preserved (except for changes beginning at 0 which are not of interest to us).

Our identifying assumption is that the number of arrest of Germans after 1933 would go in parallel to arrests of other minorities in the absence shock to German-Soviet relations conditional on our control variables (mainly the ethnicity specific time trends). Although we cannot test this assumption, we can test whether the trends were parallel prior to 1933 (pre-treatment) which could increase our confidence that they were parallel after 1933 too. This can be testing if the coefficients β_k on the lead effects are significantly different from zero.

As Bertrand et al. (2004) show, the usual standard errors are downward-biased for most DiD regressions since they do not account for the serial correlation within the units of interests (states, countries etc.). A common solution to this problem is to estimate standard errors using robust covariance matrix that allows for clustering (i.e. cluster-robust standard errors). However for small number of groups (generally less than 40), the cluster-robust standard errors are downward-biased and not reliable. Angrist and Pischke (2009, chapter 8) suggest taking the maximum of cluster-robust as a simple rule of thumb to avoid gross misjudgements in precision. More rigorous solutions are cluster bootstrapping (Cameron et al., 2008; Cameron and Miller, 2015) and using t-distribution with $G - K$ degrees of freedom (where G is number of clusters and K number of parameters) rather than the standard Normal distribution (McCaffrey and Bell, 2002; Imbens and Kolesár, 2016). Since we have small number of groups we use the generalization of McCaffrey and Bell (2002) correction to models with arbitrary sets of fixed effects by Pustejovsky and Tipton (2018).

5.2 Synthetic Control Method

However, the parallel trends assumption required for unbiasedness of difference-in-differences is often violated. Moreover, difference-in-differences often extrapolate data beyond the region of common support. To address these potential issues, we use the synthetic control method which relaxes some of the difference-in-differences assumptions and thus provides us with an useful robustness check.

Let Y_{it} be the outcome of a unit i at time t with $i = 1$ being the treated group. We denote D_{1t} as the treatment dummy, i.e. variable that equals 1 if $i = 1$ and $T > T_0$ and 0 otherwise (with T_0 being the start of the treatment). Let be Y_{1t}^N be a counterfactual outcome for the treated unit in the absence of treatment. The effect of treatment a time t, α_{1t} is assumed to be given as

$$Y_{1t} = Y_{1t}^N + \alpha_{1t} D_{1t} \tag{13}$$

Furthermore, the synthetic control method assumes that Y_{1t}^N can be expressed by the following factor model:

$$Y_{1t}^N = \delta_t + \boldsymbol{\theta}_t \boldsymbol{Z}_i + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \epsilon_{it} \tag{14}$$

where is δ_t an unknown common factor with constant factor loadings across units, $\boldsymbol{Z}_i$ is a $(1 \times r)$ vector of observed time-invariant covariates (unaffected by the treatment), $\boldsymbol{\theta}_t$ is a $(1 \times r)$ vector of unknown parameters, $\boldsymbol{\lambda}_t$ is a $(1 \times F)$ vector of unobserved time-varying factors, $\boldsymbol{\mu}_i$ is an $(F \times 1)$ vector of unknown factor loadings and ϵ_{it} is the error term with zero mean.

Notice that for constant $\boldsymbol{\lambda}_t$ for all t we get the traditional difference-in-differences model. Unlike difference-in-differences, the synthetic control method allows for unit-specific time trends as long as they can be captured by the factor model.

The synthetic control is constructed as a convex combination of available comparison units (in our case other minorities in the USSR) that most closely resembles the pre-treatment characteristics of the treated group (or more precisely, for which the average of its factor loadings $\boldsymbol{\mu}_i$ match the factor loadings of the treated unit $\boldsymbol{\mu}_1$). More formally we choose weights $W = (w_2, \ldots, w_J, w_{J+1})$ subject to $w_j \geq 0$ for $j = 1, \ldots, J, J+1$ and

$w_2 + \cdots + w_J + w_{J+1} = 1$ that minimize $\|X_1 - X_0 W\|$ where $X_1 = (Z_1, Y_1^{K_1}, \ldots, Y_1^{K_L})$ is a $(k \times 1)$ vector of pre-treatment characteristics of the treated unit and $k = r + L$ and Y^{K_l} are combinations of pre-treatment outcomes (analogously for X_0). The effect of the treatment at time t, α_{1t}, is then estimated as a difference between the outcome for synthetic control and the treated unit, i.e.:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{k=2}^{J+1} w_j^* Y_{kt}$$

6 Results

6.1 Difference-in-differences

The estimated coefficients β_k from the specification 12 are plotted in the figure 2. The coefficients between the years 1933 and 1939 (when the relations between Germany and Soviet Union were hostile) mostly range between 1 and 3 and all except one are statistically significant at 5% level. The rise of Nazism thus based on these estimated increased the arrests of Germans by the NKVD in the USSR by about 2%.

However, the pre-1933 coefficients give us some reason to doubt the validity of our model. Three of them are significantly different from 0 at 5% level and others are very close to being significant. This provide some evidence that the pre-treatment trends for German minority were not parallel with trends for other minorities. We can thus suspect that the post-treatment trends were not parallel either which would violate the basic identifying assumption of difference-in-differences. To address this concern, we apply the synthetic control method which can be valid even in the absence of parallel trends.

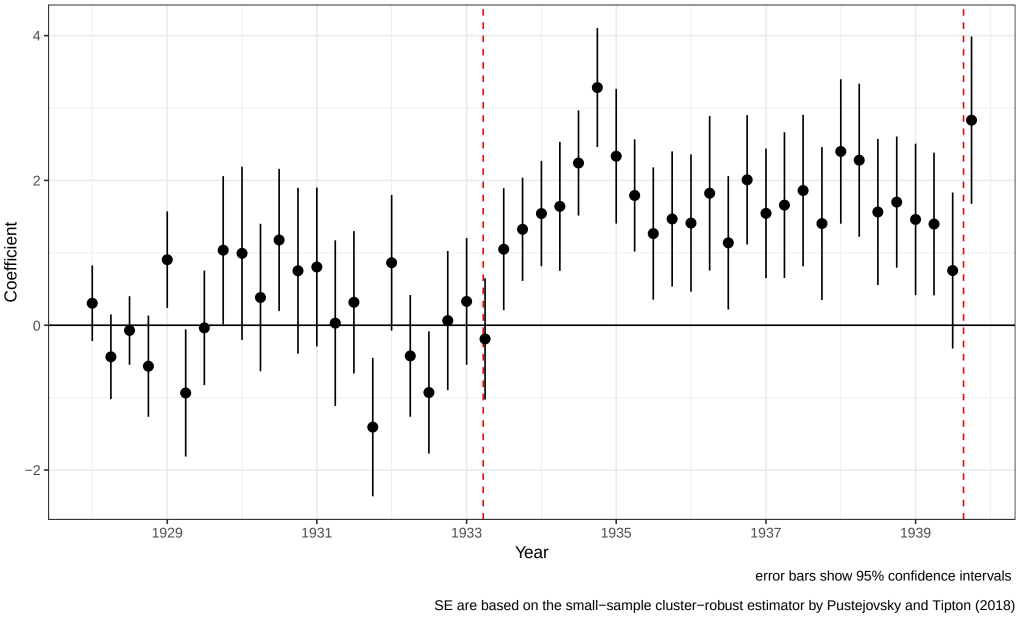


Figure 2: Estimates of coefficients β_k on the interactions of $\text{German}_i \cdot \text{YearQuart}_t^k$

6.2 Synthetic Control Method

We implemented the synthetic control method in R software using the MSCMT package (Becker and Klößner, 2018). The calculated optimal weights W of ethnic groups in the synthetic German minority are provided in the table 5 (ethnic groups with zero weight are not shown). The highest contribution in the synthetic German minority have the

Table 5: Synthetic German minority weights

Ethnic group	\$W\$-Weight
Greek	0.32
Kabardian	0.09
Chuvash	0.08
Moldovan	0.07
Belorussian	0.06
Polish	0.06
Finnish	0.06
Altai	0.04
Armenian	0.04
Georgian	0.03
Ossetian	0.02
Chechen	0.02
Kazakh	0.02
Bashkir	0.02
Jewish	0.02
Ukrainian	0.01
Bulgarian	0.01
Latvian	0.01
Chinese	0.00

Ossetians, Tatars and Pols with weights 0.39, 0.23 and 0.14 respectively. The Greek, Kabaradin, Chechen, Lithuanian and Ukrainian minorities are also represented in the synthetic control although only with very small weights.

Figure 3 shows the trends in arrests for the German minority and its synthetic control. The synthetic German minority tracks the actual values fairly well except for two large negative shocks to the actual arrests in 1931 and 1932 which the synthetic control does not capture. The trends start to diverge in the second quarter of 1933 with the actual arrests of Germans holding steady but decreasing for synthetic control. The gap between the trends for the actual German minority and its synthetic control (shown in figure 4a) keeps within the range of 1.25 to 2.5 for most of the post-1933 period. This implies that the rise of Hitler led to about 2% increase in the arrests of Germans by the NKVD in the period from 1933 to 1939. This is very similar to the estimates obtained using difference-in-differences.

The synthetic control method, however, by itself does not provide us with any measure of uncertainty of significance. This has been addressed by performing placebo tests (Abadie et al., 2010). Synthetic control method is applied iteratively to every ethnicity in the donor pool as if they were treated. By comparing the gaps from these placebo tests with the gap for the German minority, we can assess the significance of our results. Large gaps for arrests of Germans relative to other ethnic groups would suggest that the results are significant since these results would be less likely if there were no treatment effect .

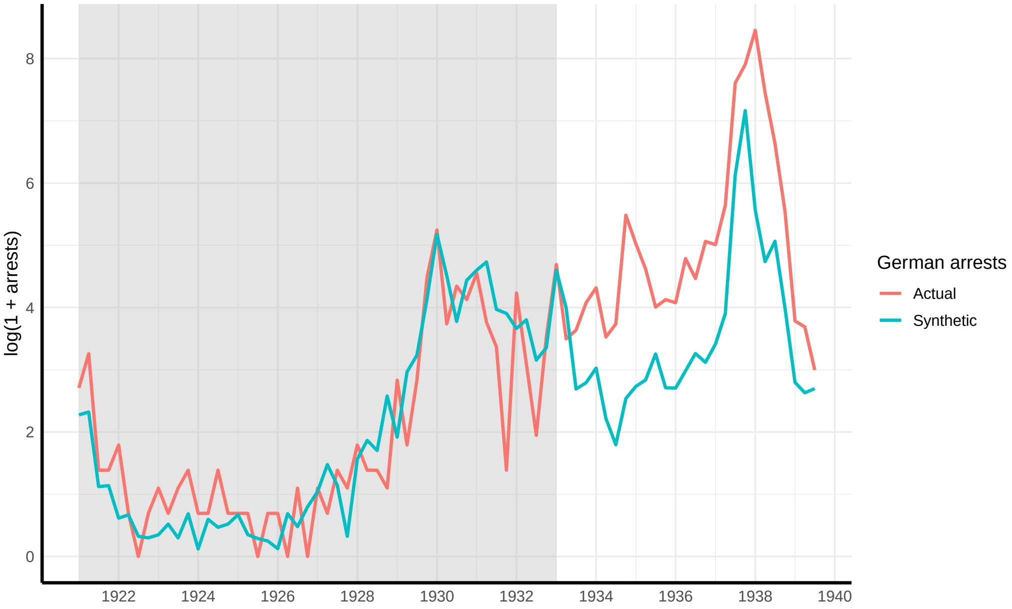
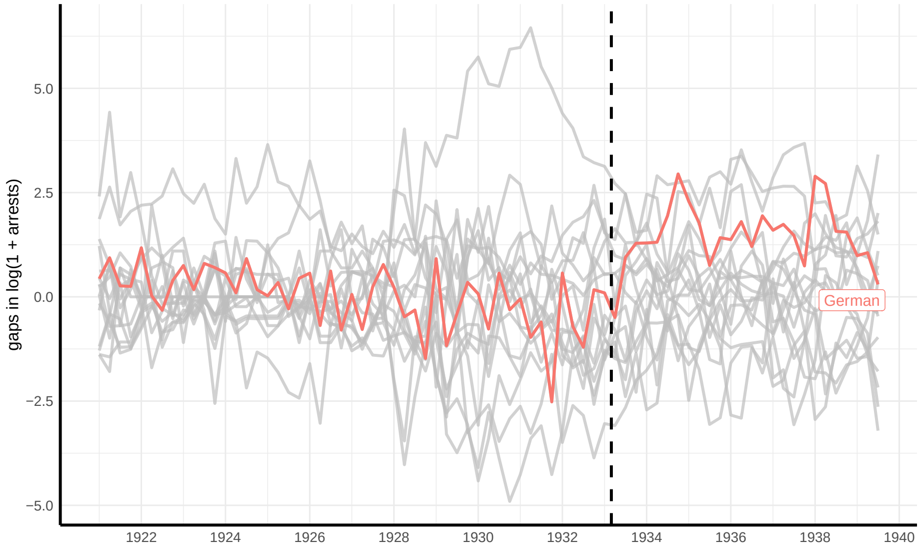


Figure 3: Comparison plot

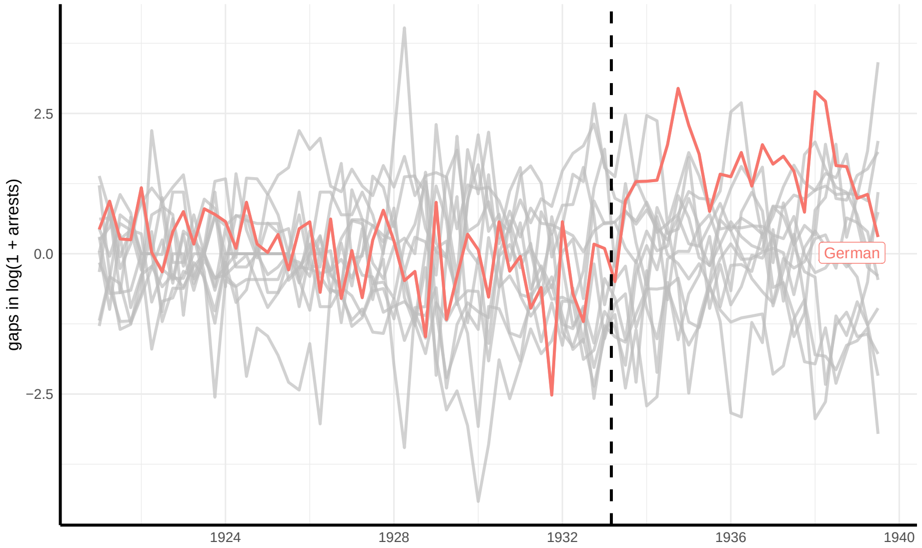
Figure 4a shows gaps between the synthetic control and the actual trends for Germans together with placebo gaps for all 17 other ethnic groups. The post-treatment gap for German minority is relatively large although not the highest. The figure also highlights that for some ethnic groups the pre-treatment gaps are large too. This indicates that synthetic control of these ethnic groups does not capture the actual pre-treatment trends well. As Abadie et al. (2010) note, placebo synthetic controls with poor pre-treatment fit do not provide good comparison for estimating rareness of large post-treatment gap for a treatment with a good pre-treatment fit. They thus recommend excluding excluding placebo groups with substantially higher pre-treatment mean squared prediction error (MSPE)

Following Abadie et al. (2010) we therefore exclude ethnic groups whose pre-treatment MSPE is 5 times higher then the same measure for German minority. This removes 4 ethnic groups with the worst pre-treatment fit. The resulting plot is shown in the figure 4b. The post-1933 gaps in German arrests now stand out more clearly.

Nevertheless, the choice of any level of the cutoff of pre-treatment MSPE is somewhat arbitrary. Alternative way to asses significance of results may be to compare the ratios of post/pre-treatment MSPE. The values of these ratios for all ethnic groups are displayed in the figure 5. Post/pre-treatment MSPE ratio for the German minority is by far the highest. The probability of German minority having the highest ratio of all under the null hypothesis of zero treatment effect is 1/17 (≈ 0.06).



(a) All ethnic groups



(b) Ethnic groups with pre-treatment MSPE higher than 5 times the MSPE of Germany excluded

Figure 4: Gaps between synthetic control and actual values for placebo tests

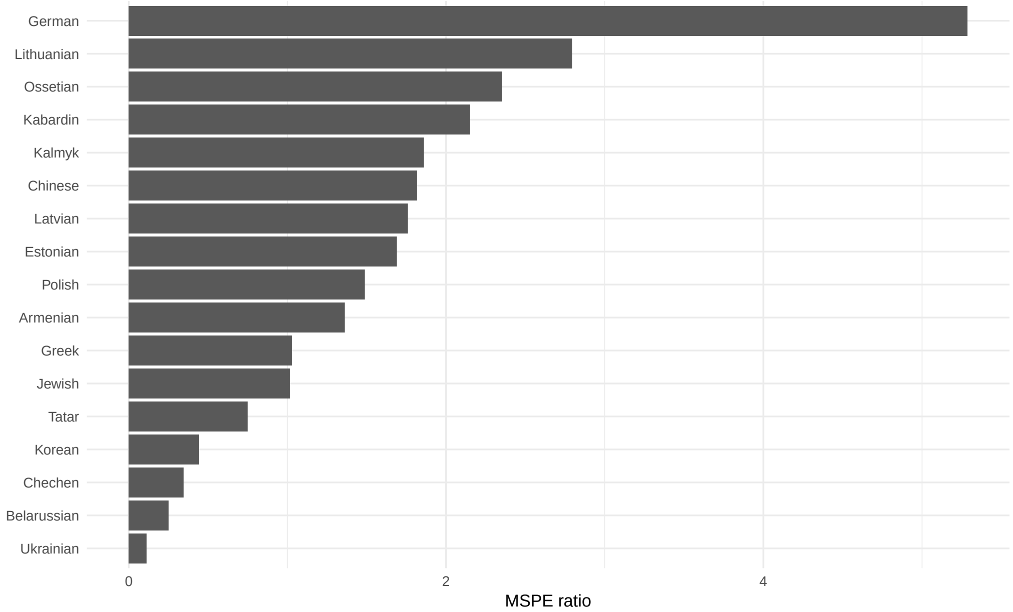


Figure 5: Ratios of post-treatment MSPE to pre-treatment MSPE

Conclusion

We used difference-in-differences and the synthetic control method to test whether the change in geopolitical relations between Soviet Union and Germany in 1933 caused the NKVD to target Soviet Germans more relative to other minority group. Both methods provide some evidence to support this hypothesis, although the estimated effects are fairly small (around 2 %). Moreover, there are several limitation to our study. The rise of Hitler might have made other non-German minorities less trustworthy in the view of the Soviet state as well because of the fear of collaboration with Germany in case of an invasion.

References

Abadie, A., Diamond, A. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program, *Journal of the American statistical Association* **105**(490): 493–505.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton.

Becker, M. and Klößner, S. (2018). Fast and reliable computation of generalized synthetic controls, *Econometrics and Statistics* **5**: 1–19.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?, *The Quarterly journal of economics* **119**(1): 249–275.

Blaydes, L. (2018). *State of Repression: Iraq under Saddam Hussein*, Princeton University Press, Princeton.

Butt, A. I. (2017). *Secession and Security: Explaining State Strategy against Separatists*, 1 edition edn, Cornell University Press, Ithaca.
URL: *https://www.jstor.org/stable/10.7591/j.ctt1w0d9w9*

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors, *The Review of Economics and Statistics* **90**(3): 414–427.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference, *Journal of Human Resources* **50**(2): 317–372.

Davenport, C. (2007). State repression and political order, *Annu. Rev. Polit. Sci.* **10**: 1–23.

Dickens, A. (2018). Ethnolinguistic favoritism in african politics, *American Economic Journal: Applied Economics* **10**(3): 370–402.

Dobson, M. (2009). *Khrushchev's Cold Summer: Gulag Returnees, Crime, and the Fate of Reform after Stalin*, Cornell University Press, Ithaca.
URL: *https://www.jstor.org/stable/10.7591/j.ctt7zdfr*

Domingos, P. and Pazzani, M. (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, *Machine Learning* **29**(2): 103–130.
URL: *https://doi.org/10.1023/A:1007413511361*

Evera, S. V. (1994). Hypotheses on Nationalism and War, *International Security* **18**(4): 5–39.

Fearon, J. D. (2003). Ethnic and cultural diversity by country, *Journal of economic growth* **8**(2): 195–222.

Gatzke, H. W. (1958). Russo-German Military Collaboration During the Weimar Republic*, *The American Historical Review* **63**(3): 565–597.
URL: *https://www.jstor.org/stable/1848881*

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, Cambridge.

Gentzkow, M., Kelly, B. T. and Taddy, M. (2019). Text as Data, *Journal of Economic Literature* (forthcoming).
URL: *https://www.aeaweb.org/articles?id=10.1257/jel.20181020from=f*

Gregory, P. R. (2009). *Terror by Quota: State Security from Lenin to Stalin*, Yale University Press, New Haven.

Greitens, S. C. (2016). *Dictators and their Secret Police: Coercive Institutions and State Violence*, Cambridge University Press, Cambridge.

Hammarström, H., Forkel, R. and Haspelmath, M. (eds) (2018). *Glottolog 3.3*, Max Planck Institute for the Science of Human History, Jena.
URL: *https://glottolog.org/*

Haslam, J. (1979). The Comintern and the Origins of the Popular Front 1934-1935, *The Historical Journal* **22**(3): 673–691.
URL: *https://www.jstor.org/stable/2638659*

Haslam, J. (1984). *Soviet Union and the Struggle for Collective Security in Europe 1933-39*, Palgrave, London.

Hofstra, B., Corten, R., van Tubergen, F. and Ellison, N. B. (2017). Sources of Segregation in Social Networks: A Novel Approach Using Facebook, *American Sociological Review* **82**(3): 625–656.
URL: *https://doi.org/10.1177/0003122417705656*

Hofstra, B. and de Schipper, N. C. (2018). Predicting ethnicity with first names in online social media networks, *Big Data & Society* **5**(1).
URL: *https://doi.org/10.1177/2053951718761141*

Imbens, G. W. and Kolesár, M. (2016). Robust Standard Errors in Small Samples: Some Practical Advice, *The Review of Economics and Statistics* **98**(4): 701–712.
URL: *https://www.mitpressjournals.org/doi/10.1162/*$REST_{a0}0552$

Kotkin, S. (2017). *Stalin: Waiting for Hitler, 1929-1941*, Penguin Press, London.

Kravchenko, V. (1947). *I Chose Freedom: The Personal and Political Life of a Soviet Official*, Robert Hale Limited, London.

Lupu, N. and Peisakhin, L. (2017). The legacy of political violence across generations, *American Journal of Political Science* **61**(4): 836–851.

Martin, T. (1998). The Origins of Soviet Ethnic Cleansing, *The Journal of Modern History* **70**(4): 813–861.
URL: *https://www.jstor.org/stable/10.1086/235168*

Martin, T. D. (2001). *The Affirmative Action Empire: Nations and Nationalism in the Soviet Union, 1923-1939*, Cornell University Press, Ithaca.

Mateos, P. (2007). A review of name-based ethnicity classification methods and their potential in population studies, *Population, Space and Place* **13**(4): 243–263.

McCaffrey, D. F. and Bell, R. M. (2002). Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples, *Survey Methodology* **28**(2): 169–181.

McNamee, L. and Zhang, A. (2019). Demographic Engineering and International Conflict: Evidence from China and the Former USSR, *International Organization* (forthcoming).

Memorial (2017). Zhertvy politicheskogo terrora v SSSR [victims of political terror in the USSR].
URL: *http://base.memo.ru/*

Morgan, R. P. (1963). The Political Significance of German-Soviet Trade Negotiations, 1922-5, *The Historical Journal* **6**(2): 253–271.
URL: *https://www.jstor.org/stable/3020490*

Mylonas, H. (2013). *The Politics of Nation-Building: Making Co-Nationals, Refugees, and Minorities*, Cambridge University Press, Cambridge.

Polian, P. (2003). *Against Their Will: The History and Geography of Forced Migrations in the USSR*, Central European University Press, Budapest.

Pustejovsky, J. E. and Tipton, E. (2018). Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models, *Journal of Business*

& Economic Statistics **36**(4): 672–683.
URL: *https://doi.org/10.1080/07350015.2016.1247004*

Robinson, R. and Slevin, J. (1988). *Black on Red: My 44 Years Inside the Soviet Union*, Acropolis Books, Washington, D.C.

Rozenas, A., Schutte, S. and Zhukov, Y. (2017). The political legacy of violence: The long-term impact of Stalin's repression in Ukraine, *The Journal of Politics* **79**(4): 1147–1161.

Snyder, T. (2011). *Bloodlands: Europe between Hitler and Stalin*, Vintage, London.

Svolik, M. W. (2012). *The Politics of Authoritarian Rule*, Cambridge University Press, Cambridge.

Toft, M. D. (2005). *The Geography of Ethnic Violence: Identity, Interests, and the Indivisibility of Territory*, Princeton University Press, Princeton.

Tooze, A. (2008). *The Wages of Destruction: The Making and Breaking of the Nazi Economy*, Penguin Books, New York.

Walter, B. F. (2009). *Reputation and Civil War: Why Separatist Conflicts Are So Violent*, 1 edition edn, Cambridge University Press, Cambridge.

Weinberg, G. L. (2010). *Hitler's Foreign Policy 1933-1939: The Road to World War II*, Enigma Books.

Wintrobe, R. (1998). *The Political Economy of Dictatorship*, Cambridge University Press, Cambridge.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*, Nelson Education.

Zhukov, Y. M. and Talibova, R. (2018). Stalin's terror and the long-term political effects of mass repression, *Journal of Peace Research* **55**(2): 267–283.

List of tables and figures

List of Tables

List of Figures

Appendix

Additional Tables

Table 6: Total arrest by ethnicity, 1921-1960

	Reference		
Ethnicity	Only Labeled	Labeled + Unadj. Imputation	Labeled + Adj. Imputation
Russian	550 280	1 064 596	1 069 379
Belorussian	67 613	85 517	72 979
Polish	61 221	85 258	79 742
German	60 798	168 419	169 955
Ukrainian	54 403	91 814	97 042
Kazakh	37 125	46 540	43 541
Tatar	32 095	72 417	71 351
Jewish	31 050	43 710	42 613
Latvian	15 444	21 628	18 796
Chinese	9 693	11 506	10 466
Estonian	9 402	15 561	13 380
Chuvash	8 910	14 930	26 520
Bashkir	8 428	17 876	18 615
Finnish	8 337	14 594	13 550
Mordvin	6 011	12 682	20 642
Buryat	5 679	6 735	6 715
Mari	5 383	7 482	12 288
Lithuanian	4 651	5 474	5 522
Karelian	4 174	9 941	5 379
Korean	4 060	8 821	11 560
Komi	3 613	5 834	4 281
Ossetian	3 237	3 724	3 419
Udmurt	3 082	4 454	5 566
Armenian	2 937	4 850	4 674
Kabardian	2 733	4 438	4 021
Greek	2 246	24 500	25 514
Khakas	2 221	8 137	6 136
Altai	1 894	2 477	2 471
Georgian	1 621	3 049	1 993
Yakut	1 544	2 909	1 572
Moldovan	1 392	2 765	2 719
Kalmyk	1 293	2 168	2 059
Japanese	1 231	14 571	10 821
Uzbek	1 061	4 044	7 470
Hungarian	1 018	1 611	1 119
Bulgarian	1 015	2 479	1 904
Balkar	861	4 740	3 423
Chechen	696	8 508	11 548

Table 7: Naive Bayes Performance Measures by Ethnicity

Ethnicity	Sensitivity	Specificity
Altai	0.483	1.000
Armenian	0.796	0.999
Balkar	0.971	0.999
Bashkir	0.481	0.997
Belorussian	0.502	0.976
Bulgarian	0.402	0.999
Buryat	0.774	1.000
Estonian	0.702	0.996
Finnish	0.789	0.997
Georgian	0.569	0.999
German	0.875	0.988
Greek	0.701	0.994
Hungarian	0.379	0.998
Chechen	0.590	0.998
Chinese	0.915	0.997
Chuvash	0.100	0.995
Japanese	0.968	0.996
Jewish	0.864	0.997
Kabardian	0.882	0.999
Kalmyk	0.848	1.000
Karelian	0.173	0.991
Kazakh	0.825	0.999
Khakas	0.832	0.997
Komi	0.248	0.998
Korean	0.495	0.999
Latvian	0.673	0.995
Lithuanian	0.572	0.999
Mari	0.196	0.999
Moldovan	0.285	0.999
Mordvin	0.172	0.996
Ossetian	0.836	1.000
Polish	0.786	0.980
Russian	0.876	0.875
Tatar	0.809	0.995
Udmurt	0.092	0.999
Ukrainian	0.423	0.977
Uzbek	0.330	0.999
Yakut	0.199	0.997

Table 8: Confusion Matrix (based on 10-f

	Reference																	
Prediction	Altai	Armenian	Balkar	Bashkir	Belorussian	Bulgarian	Buryat	Estonian	Finnish	Georgian	German	Greek	Hungarian	Chechen	Chinese	Chuvash	Japanese	Jewish	Kabar
Altai	972	1	5	2	13	1	9	2	3	2	11	4	1	2	68	4	0	15	
Armenian	2	2688	12	1	41	2	5	6	5	35	43	11	3	0	0	3	1	79	
Balkar	3	2	48517	22	2	6	0	0	0	6	25	1	0	70	0	1	0	11	
Bashkir	0	2	30	4075	0	0	2	0	0	0	1	0	0	16	2	10	0	2	
Belorussian	10	28	5	0	36631	14	16	92	53	39	424	54	48	0	11	44	1	448	
Bulgarian	4	11	5	1	79	414	23	4	10	18	50	49	4	0	1	8	1	26	
Buryat	12	0	8	1	5	1	4402	0	0	2	1	0	0	6	3	3	0	9	
Estonian	3	7	1	0	223	5	6	6681	276	11	1022	12	53	0	8	8	0	171	
Finnish	7	11	6	1	198	3	10	177	7183	12	326	9	16	0	2	12	0	131	
Georgian	4	43	12	0	113	9	13	14	24	975	75	41	2	0	0	5	0	120	
German	12	41	37	0	590	23	5	802	345	47	68862	57	78	0	9	40	0	929	
Greek	21	41	16	0	389	76	9	38	41	104	301	2332	5	0	1	42	2	145	
Hungarian	4	23	4	0	130	10	19	50	76	20	266	5	389	2	1	2	0	161	
Chechen	43	56	218	78	14	8	98	6	22	14	44	10	6	562	4	8	0	208	
Chinese	2	1	0	3	8	2	9	7	1	3	28	0	1	1	8938	1	9	3	
Chuvash	26	23	3	1	515	22	64	9	24	19	34	26	7	0	9	898	0	83	
Japanese	226	5	30	33	4	2	127	5	19	3	363	2	4	95	103	3	1192	25	
Jewish	2	22	3	1	195	1	1	57	53	31	516	14	30	0	5	6	1	37949	
Kabardian	3	3	625	11	2	1	3	1	1	4	11	0	1	26	1	0	0	26	
Kalmyk	9	3	31	2	3	0	68	0	0	0	4	1	0	8	0	0	1	11	
Karelian	33	16	1	2	821	35	35	84	166	22	310	53	35	0	60	108	1	66	
Kazakh	59	3	72	129	0	0	6	0	0	4	5	1	1	49	17	0	0	3	
Khakas	14	6	10	1	162	9	24	11	19	23	77	17	4	1	4	29	0	70	
Komi	9	12	1	3	196	10	17	15	15	13	45	17	10	1	4	33	0	14	
Korean	16	8	5	0	42	2	4	9	3	4	94	6	0	2	196	6	8	20	
Latvian	4	8	2	0	351	11	3	770	218	9	1258	12	52	0	7	13	1	155	
Lithuanian	0	2	1	0	66	0	4	16	12	1	40	0	5	0	0	0	0	42	
Mari	16	5	7	27	32	1	18	0	3	1	6	2	0	1	8	43	1	10	
Moldovan	13	29	2	0	107	11	20	3	13	34	38	23	2	0	2	11	0	33	
Mordvin	14	14	6	2	382	6	28	15	21	15	43	17	7	0	4	83	0	43	
Ossetian	7	13	124	2	19	0	30	2	3	15	5	2	0	12	2	1	0	35	
Polish	1	12	8	0	10265	10	1	105	48	17	1301	29	113	0	5	10	3	643	
Russian	379	172	48	27	16658	262	491	480	384	120	2670	383	113	5	239	7383	4	1711	
Tatar	11	1	75	3990	1	1	9	0	1	2	26	3	0	68	20	25	1	14	
Udmurt	9	9	4	9	85	4	16	4	9	9	15	5	2	1	0	24	1	32	
Ukrainian	29	36	7	1	4438	37	10	40	19	51	310	90	29	0	8	78	1	400	
Uzbek	5	0	33	45	0	0	2	0	0	0	0	0	0	25	13	1	0	6	
Yakut	30	19	3	0	221	30	83	15	32	29	64	39	5	0	14	44	2	51	

Table 9: Confusion Matrix (based on 10-fol

	Reference																	
Prediction	Altai	Armenian	Balkar	Bashkir	Belorussian	Bulgarian	Buryat	Estonian	Finnish	Georgian	German	Greek	Hungarian	Chechen	Chinese	Chuvash	Japanese	Jewish	Kabar
Altai	0.483	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.001	0.000	0.001	0.001	0.002	0.007	0.000	0.000	0.000	0
Armenian	0.001	0.796	0.000	0.000	0.001	0.002	0.001	0.001	0.001	0.020	0.001	0.003	0.003	0.000	0.000	0.000	0.001	0.002	0
Balkar	0.001	0.001	0.971	0.003	0.000	0.006	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.073	0.000	0.000	0.000	0.000	0
Bashkir	0.000	0.001	0.001	0.481	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.001	0.000	0.000	0
Belorussian	0.005	0.008	0.000	0.000	0.502	0.014	0.003	0.010	0.006	0.023	0.005	0.016	0.047	0.000	0.001	0.005	0.001	0.010	0
Bulgarian	0.002	0.003	0.000	0.000	0.001	0.402	0.004	0.000	0.001	0.011	0.001	0.015	0.004	0.000	0.000	0.001	0.001	0.001	0
Buryat	0.006	0.000	0.000	0.000	0.000	0.001	0.774	0.000	0.000	0.001	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0
Estonian	0.001	0.002	0.000	0.000	0.003	0.005	0.001	0.702	0.030	0.006	0.013	0.004	0.052	0.000	0.001	0.001	0.000	0.004	0
Finnish	0.003	0.003	0.000	0.000	0.003	0.003	0.002	0.019	0.789	0.007	0.004	0.003	0.016	0.000	0.000	0.001	0.000	0.003	0
Georgian	0.002	0.013	0.000	0.000	0.002	0.009	0.002	0.001	0.003	0.569	0.001	0.012	0.002	0.000	0.000	0.001	0.000	0.003	0
German	0.006	0.012	0.001	0.000	0.008	0.022	0.001	0.084	0.038	0.027	0.875	0.017	0.076	0.000	0.001	0.004	0.000	0.021	0
Greek	0.010	0.012	0.000	0.000	0.005	0.074	0.002	0.004	0.005	0.061	0.004	0.701	0.005	0.000	0.000	0.005	0.002	0.003	0
Hungarian	0.002	0.007	0.000	0.000	0.002	0.010	0.003	0.005	0.008	0.012	0.003	0.002	0.379	0.002	0.000	0.000	0.000	0.004	0
Chechen	0.021	0.017	0.004	0.009	0.000	0.008	0.017	0.001	0.002	0.008	0.001	0.003	0.006	0.590	0.000	0.001	0.000	0.005	0
Chinese	0.001	0.000	0.000	0.000	0.000	0.002	0.002	0.001	0.000	0.002	0.000	0.000	0.001	0.001	0.915	0.000	0.007	0.000	0
Chuvash	0.013	0.007	0.000	0.000	0.007	0.021	0.011	0.001	0.003	0.011	0.000	0.008	0.007	0.000	0.001	0.100	0.000	0.002	0
Japanese	0.112	0.001	0.001	0.004	0.000	0.002	0.022	0.001	0.002	0.002	0.005	0.001	0.004	0.100	0.011	0.000	0.968	0.001	0
Jewish	0.001	0.007	0.000	0.000	0.003	0.001	0.000	0.006	0.006	0.018	0.007	0.004	0.029	0.000	0.001	0.001	0.001	0.864	0
Kabardian	0.001	0.001	0.013	0.001	0.000	0.001	0.001	0.000	0.000	0.002	0.000	0.000	0.001	0.027	0.000	0.000	0.000	0.001	0
Kalmyk	0.004	0.001	0.001	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.001	0.000	0
Karelian	0.016	0.005	0.000	0.000	0.011	0.034	0.006	0.009	0.018	0.013	0.004	0.016	0.034	0.000	0.006	0.012	0.001	0.002	0
Kazakh	0.029	0.001	0.001	0.015	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.001	0.051	0.002	0.000	0.000	0.000	0
Khakas	0.007	0.002	0.000	0.000	0.002	0.009	0.004	0.001	0.002	0.013	0.001	0.005	0.004	0.001	0.000	0.003	0.000	0.002	0
Komi	0.004	0.004	0.000	0.000	0.003	0.010	0.003	0.002	0.002	0.008	0.001	0.005	0.010	0.001	0.000	0.004	0.000	0.000	0
Korean	0.008	0.002	0.000	0.000	0.001	0.002	0.001	0.001	0.000	0.002	0.001	0.002	0.000	0.002	0.020	0.001	0.006	0.000	0
Latvian	0.002	0.002	0.000	0.000	0.005	0.011	0.001	0.081	0.024	0.005	0.016	0.004	0.051	0.000	0.001	0.001	0.001	0.004	0
Lithuanian	0.000	0.001	0.000	0.000	0.001	0.000	0.001	0.002	0.001	0.001	0.001	0.000	0.005	0.000	0.000	0.000	0.000	0.001	0
Mari	0.008	0.001	0.000	0.003	0.000	0.001	0.003	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.001	0.005	0.001	0.000	0
Moldovan	0.006	0.009	0.000	0.000	0.001	0.011	0.004	0.000	0.001	0.020	0.000	0.007	0.002	0.000	0.000	0.001	0.000	0.001	0
Mordvin	0.007	0.004	0.000	0.000	0.005	0.006	0.005	0.002	0.002	0.009	0.001	0.005	0.007	0.000	0.000	0.009	0.000	0.001	0
Ossetian	0.003	0.004	0.002	0.000	0.000	0.000	0.005	0.000	0.000	0.009	0.000	0.001	0.000	0.013	0.000	0.000	0.000	0.001	0
Polish	0.000	0.004	0.000	0.000	0.141	0.010	0.000	0.011	0.005	0.010	0.017	0.009	0.110	0.000	0.001	0.001	0.002	0.015	0
Russian	0.188	0.051	0.001	0.003	0.228	0.255	0.086	0.050	0.042	0.070	0.034	0.115	0.110	0.005	0.024	0.821	0.003	0.039	0
Tatar	0.005	0.000	0.002	0.471	0.000	0.001	0.002	0.000	0.000	0.001	0.000	0.001	0.000	0.071	0.002	0.003	0.001	0.000	0
Udmurt	0.004	0.003	0.000	0.001	0.001	0.004	0.003	0.000	0.001	0.005	0.000	0.002	0.002	0.001	0.000	0.003	0.001	0.001	0
Ukrainian	0.014	0.011	0.000	0.000	0.061	0.036	0.002	0.004	0.002	0.030	0.004	0.027	0.028	0.000	0.001	0.009	0.001	0.009	0
Uzbek	0.002	0.000	0.001	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.001	0.000	0.000	0.000	0
Yakut	0.015	0.006	0.000	0.000	0.003	0.029	0.015	0.002	0.004	0.017	0.001	0.012	0.005	0.000	0.001	0.005	0.002	0.001	0

Table 10: Arrest Date Imputation - Model Results

	Dependent variable:	
	I^y	$log(y^{\text{pos}})$
	logistic	*OLS*
	(1)	(2)
(Intercept)	−0.771*** (0.161)	0.888*** (0.025)
Year of Process - 1922	1.630*** (0.586)	1.038*** (0.033)
Year of Process - 1923	0.955* (0.510)	0.976*** (0.039)
Year of Process - 1924	2.128** (1.004)	1.122*** (0.043)
Year of Process - 1925	1.019* (0.586)	1.112*** (0.043)
Year of Process - 1926	0.508* (0.284)	0.972*** (0.027)
Year of Process - 1927	0.346 (0.234)	0.904*** (0.024)
Year of Process - 1928	−0.260** (0.123)	0.422*** (0.016)
Year of Process - 1929	−0.209** (0.101)	0.307*** (0.012)
Year of Process - 1930	0.012 (0.103)	0.754*** (0.012)
Year of Process - 1931	0.166 (0.111)	0.695*** (0.013)
Year of Process - 1932	0.299*** (0.106)	0.463*** (0.012)
Year of Process - 1933	0.193 (0.139)	0.457*** (0.016)
Year of Process - 1934	−0.198* (0.116)	0.602*** (0.014)
Year of Process - 1935	−0.206* (0.112)	0.874*** (0.014)
Year of Process - 1936	0.858*** (0.099)	−0.298*** (0.011)
Year of Process - 1937	1.116*** (0.101)	0.486*** (0.012)
Year of Process - 1938	1.845*** (0.151)	2.010*** (0.013)
Year of Process - 1939	1.560*** (0.162)	1.579*** (0.013)
Year of Process - 1940	0.463*** (0.113)	0.705*** (0.013)
Year of Process - 1941	0.282*** (0.109)	0.641*** (0.013)
Year of Process - 1942	0.234** (0.114)	0.833*** (0.013)
Year of Process - 1943	−0.422*** (0.118)	0.804*** (0.015)
Year of Process - 1944	0.175 (0.127)	0.936*** (0.015)
Year of Process - 1945	0.264* (0.142)	1.182*** (0.016)
Year of Process - 1946	0.164 (0.161)	0.987*** (0.018)
Year of Process - 1947	0.231 (0.179)	0.860*** (0.020)
Year of Process - 1948	0.810*** (0.192)	0.735*** (0.017)
Year of Process - 1949	0.512*** (0.186)	0.953*** (0.019)
Year of Process - 1950	0.532*** (0.188)	0.908*** (0.019)
Year of Process - 1951	−0.080 (0.197)	0.844*** (0.024)
Year of Process - 1952	0.077 (0.269)	0.619*** (0.031)
Year of Process - 1953	0.003 (0.589)	1.680*** (0.070)
Year of Process - 1954	−0.526 (0.462)	2.253*** (0.071)
Year of Process - 1955	−0.713* (0.425)	1.324*** (0.071)
Year of Process - 1956	0.950*** (0.367)	0.683*** (0.029)
Year of Process - 1957	0.595 (0.367)	0.813*** (0.034)
Year of Process - 1958	−0.232 (0.333)	1.036*** (0.044)
Year of Process - 1959	−0.716 (0.516)	1.042*** (0.087)
Year of Process - 1960	4.292*** (0.094)	3.697*** (0.011)
Observations	812,592	805,800
R^2		0.235
Adjusted R^2		0.235
Log Likelihood	−38,300.970	
Akaike Inf. Crit.	76,681.930	
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 11: Pre-treatment characteristics of ethnic groups in the USSR

Ethnic group	Total population	Ling. similarity to Russian	Urbanization rate	Ind. country
Armenian	1 567 568	1	35.45	0
Belorussian	4 738 923	4	10.32	0
Estonian	154 666	0	23.00	1
German	1 238 549	1	14.92	1
Greek	213 765	1	21.21	1
Chechen	318 522	0	0.98	0
Chinese	10 247	0	64.87	1
Jewish	2 599 973	1	82.43	0
Kabardian	139 925	0	1.27	0
Kalmyk	129 321	0	1.29	0
Korean	86 999	0	10.52	0
Latvian	141 703	2	42.31	1
Lithuanian	41 463	2	63.16	1
Ossetian	272 272	1	7.86	0
Polish	782 334	3	32.75	1
Tatar	2 916 536	0	15.48	0
Ukrainian	31 194 976	4	10.54	0
Altai	39 062	0	0.30	0
Balkar	33 307	0	1.23	0
Bashkir	713 693	0	2.12	0
Bulgarian	111 296	3	6.26	0
Buryat	237 501	0	1.05	0
Finnish	134 701	0	10.55	1
Georgian	1 821 184	0	16.93	0
Hungarian	5 476	0	63.33	1
Chuvash	1 117 419	0	1.60	0
Japanese	93	0	76.34	1
Karelian	248 120	0	2.91	0
Kazakh	3 968 289	0	2.18	0
Khakas	45 608	0	1.08	0
Komi	375 871	0	2.56	0
Mari	428 192	0	0.84	0
Moldovan	278 905	1	4.86	0
Mordvin	1 340 415	0	2.19	0
Russian	77 791 124	5	21.32	1
Udmurt	504 187	0	1.21	0
Uzbek	3 904 622	0	18.66	0
Yakut	240 709	0	2.20	0

Note:

Total population and urbanization rate of the ethnic group in the USSR is taken from 1926 census. The linguistic similarity to Russian is measured by the number of common nodes in the language tree (cladistic similarity). Independent country equals one if the ethnic group was a core group in an independent state that existed in the interwar period.

Additional Figures

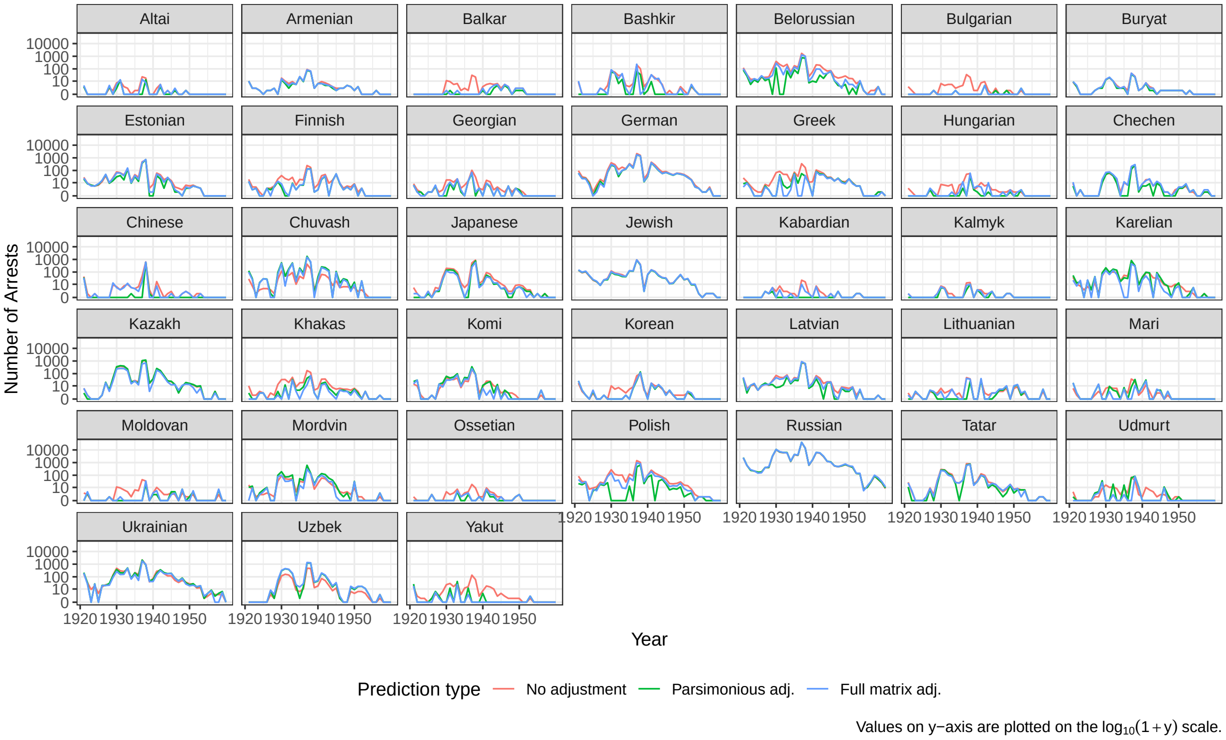


Figure 6: Number of Predicted Arrests by Ethnicity, Year, and Prediction Adjustment

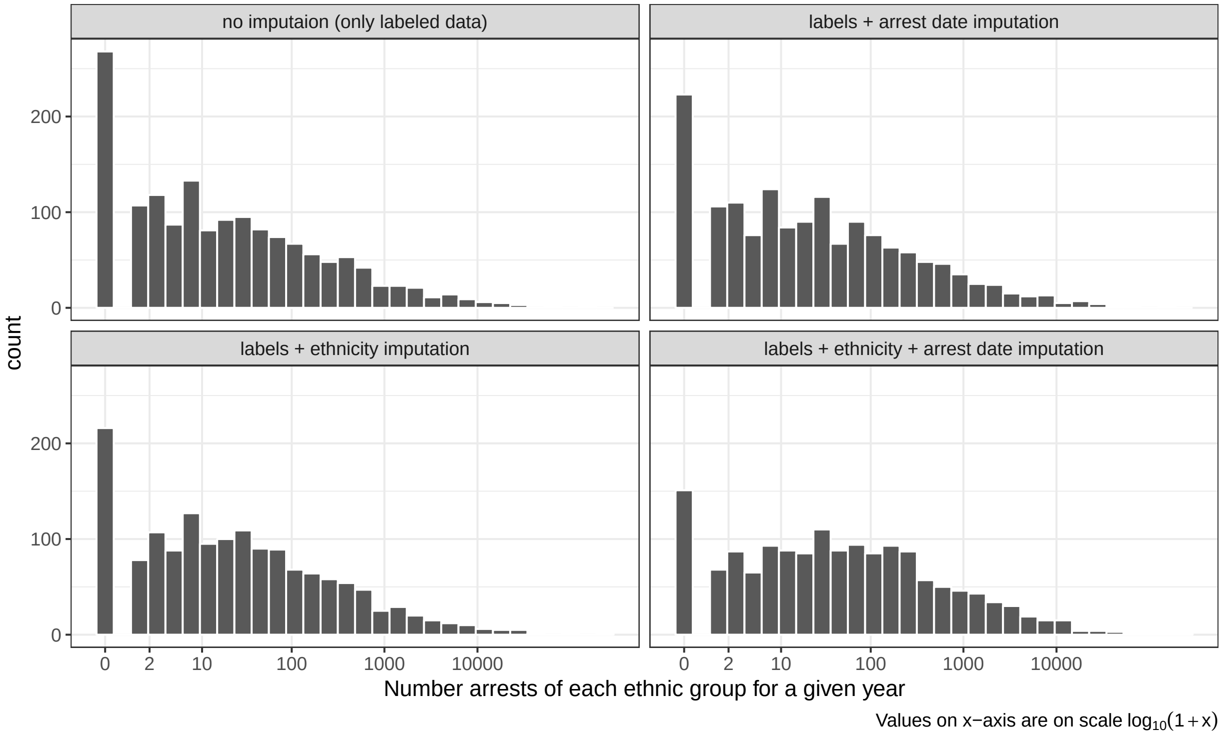


Figure 7: Histograms of Arrests by Ethnicity and Year

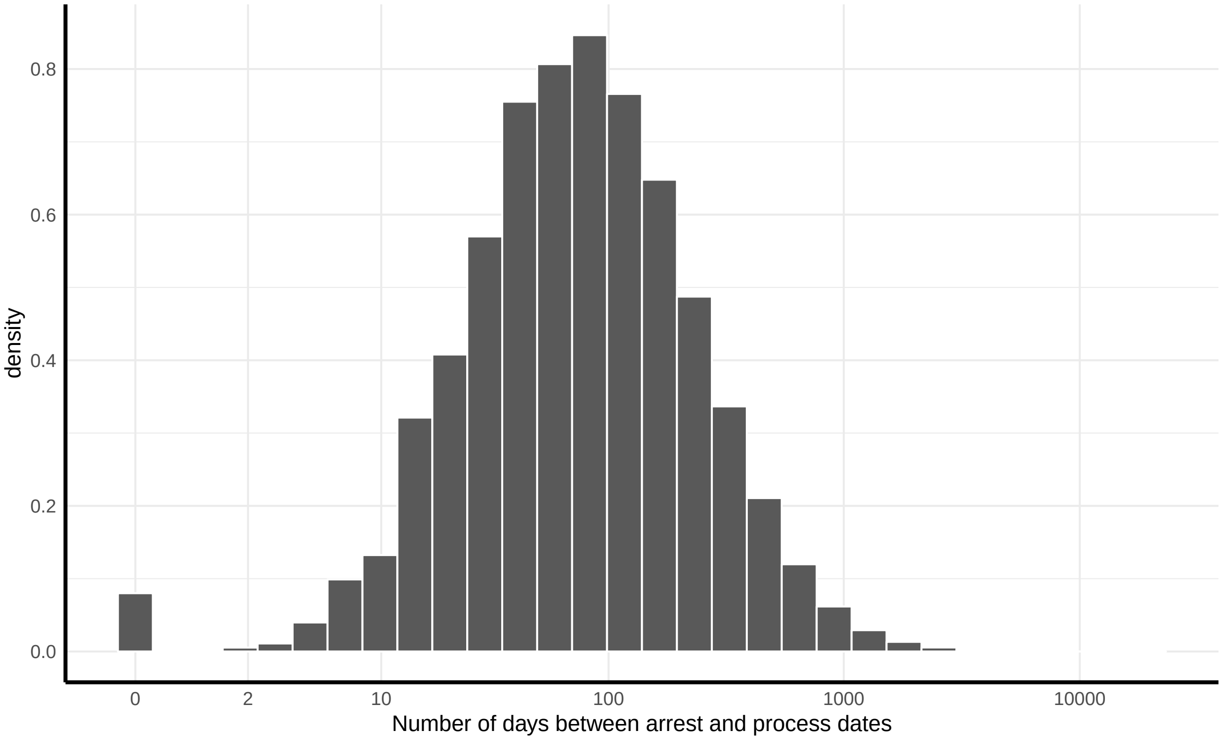
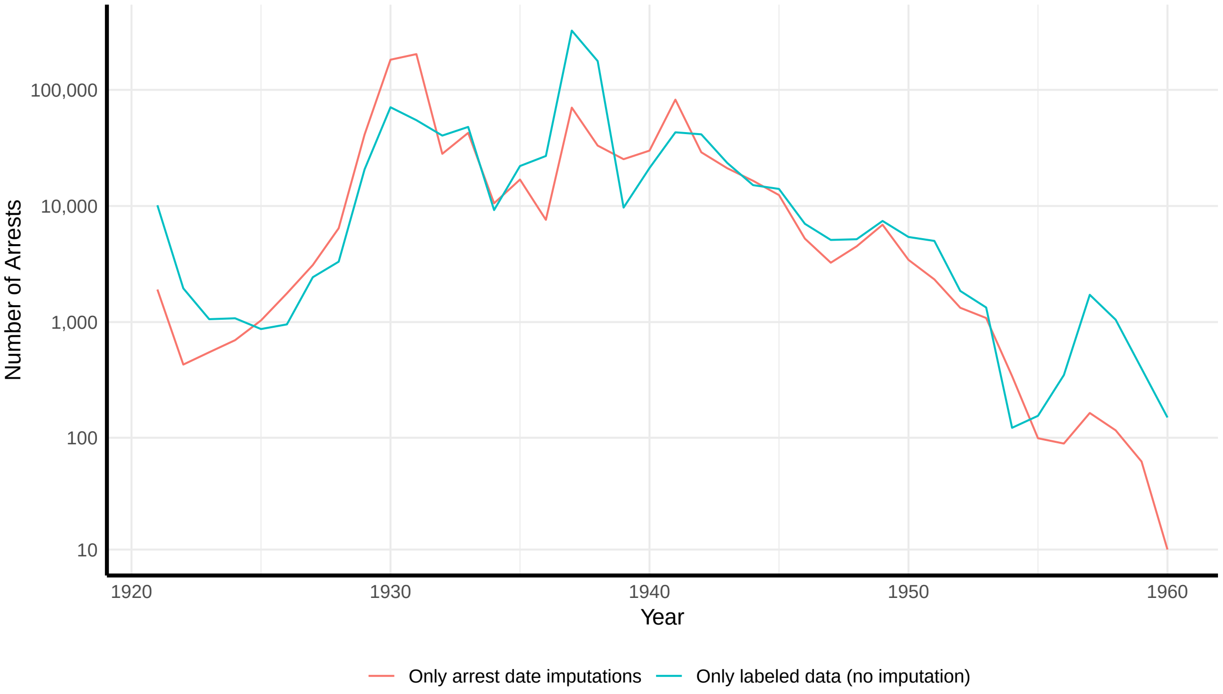


Figure 8: Histogram of Imputed Number of Days between Arrest and Process



Values on y−axis are plotted on the $\log_{10}(1+y)$ scale.

Figure 9: Time Series of Arrests