The Geopolitics of Repressions

Martin Kosík

March 16, 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 2.3 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.

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 (the NKVD) 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; Svolik, 2012; Greitens, 2016; Blaydes, 2018) with less attention being given to external forces.

As was mentioned, Mylonas (2013) proposes a theory how of geopolitical relations influence the attitude of a state towards its minorities. He also 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.

According to Blaydes (2018), a state will resort to collective punishment (based on ethnicity, religion or community membership) if it faces environment with highly asymmetric information in which it cannot identify the likely transgressors. The logic behind this is that the members of the community will police its members to avoid collective punishment.

McNamee and Zhang (n.d.) 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

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 which they committed to co-operate for defense against communistic disintegration. Moreover, the pact

Meanwhile in the Soviet Union, many people were persecuted for alleged cooperations 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 ideological enemies caused great shock and astonishment both among official and ordinary people.

Except the brief period of limited cooperation, the Germany represented an ideological and geopolitical opponent. The Soviet propaganda portrayed Nazi Germany as an existential enemy and rank-and-file NKVD officers would perceive it as such (which is why Molotov-Ribbentrop pact caused such a surprise). (Kotkin, 2017)

The pact was surprising for many people including Party officials. Robinson, Black on Red, 137. "It left us all stunned, bewildered, and groggy with disbelief," recalled one loyal party member (who later defected). 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.

Other

2.2 Political repressions in the Soviet Union

2.2.1 Ethnic repressions in the Soviet Union

In the 1920s, the Soviet policy towards its ethnic minorities was largely accommodating (Martin, 2001). The languages and culture of minorities were even often promoted and minorities were encouraged to enter local goverments and party structures (so-called *korenizatsiya* policy). 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. The Soviet state then gradually began to target ethnic minorities for repressions which culminated in the mass national operations of the NKVD of 1937-1938 resulting in more than 100 000 people being killed and many more sent to the Gulags (forced labor camps) (Martin, 1998; Gregory, 2009; Snyder, 2011). The persecutions further escalated with the World War II. Following the German invasion into the Soviet Union in 1941, Stalin ordered deportation of about 400 000 Volga Germans into Kazakhstan and Siberia (Polian, 2003).

3 Data

Our data on soviet repressions come from Zhukov and Talibova (2018)[1] who use the Victims of Political Terror archive [2] collected by a Russian NGO Memorial. 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. The Memorial archives include 2.3 million individual arrests by the Soviet secret police (NKVD) between the years 1921 and 1959 with names of each person, date of arrest, the place of birth for all observations and in many cases additional information such as ethnicity, occupation and party membership. However the data are not complete and include about 70% of estimated 3.8 million convicted for political reasons.

We created our main dataset by counting number of arrest for each ethnicity by yearquater. A few people who were categorized as having multiple ethnicities were dropped from the dataset and not counted. With 17 minorities (Armenian, Belarussian, Estonian,

[1]In particular, we downloaded the data from the replications file archive of the journal available at https://www.prio.org/JPR/Datasets/

[2]The Memorial archive can be accessed at http://base.memo.ru/ (new version) or at http://lists.memo.ru/ (older version)

German, Greek, Chechen, Chinese, Jewish, Kabardin, Kalmyk, Korean, Latvian, Lithuanian, Ossetian, Polish, Tatar and Ukrainian) and 148 time periods (from 1921 to 1958) this gives us 2652 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 5, both in the appendix.

In addition to data on repressions, we also obtained some information on a few characteristics of the 17 ethnic groups in the USSR. Specifically, we acquired total population of the ethnic groups and their urbanization rate from 1926 Soviet Census from the Demoscope website.[3] 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 6 in the appendix.

3.1 Predicting Ethnicity from Names

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

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

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

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

All the terms in this equation 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

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

It is important to note that the conditional independence assumption often does not hold in the data. The estimated posterior probabilities 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

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

this respect, Naive Bayes classifier have been shown to perform well in many applications, despite its often violated assumptions (Domingos and Pazzani, 1997).

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 3. 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

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

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

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

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. We can im

$$P_{it} = \sum_{j=1}^{18} b_{ij} R_{jt} \qquad i = 1, \ldots, 18$$

First we ap

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

Again, we can easily calculate th

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

We get bigger picture by examining the confusion matrix provided in the table 4 in the appendix. The confusion matrix is defined as

$$A = (a_{ij})_{i=1,\ldots,18,\, j=1,\ldots,18}$$

where a_{ij} is number of people with ethnicity j that were classified as belonging to ethnic group i. We can adjust the matrix as

$$B = \left(\frac{a_{ij}}{\sum_j a_{ij}} \right)_{i=1,\ldots,18,\, j=1,\ldots,18}$$

$$\begin{pmatrix} a_1t & a_{12} & \ldots & a_{1n} \\ a_{21} & a_2 & \ldots & a_{2n} \\ & \cdots\cdots & & \\ a_{n1} & a_{n2} & \ldots & D_nt \end{pmatrix}$$

4 Difference-in-differences

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

$$\log\left(1+y_{it}\right) = \sum_{k=1928\text{Q1}}^{1939\text{Q4}} \beta_k \,\text{German}_i \cdot \text{YearQuart}_t^k + \lambda_t + a_i + E_i \cdot t \; + E_i \cdot t^2 \; + u_{it} \quad (1)$$

where y_{it} is number of arrests of people with ethnicity i in year-quarter t, λ is year fixed effect, a is ethnicity fixed effect (both captured by respective dummy variables) and YearQuart_t^k are dummy variables that equals 1 if its year-quarter k equals to t and 0 otherwise (except for 1939Q4 which is equal to 1 after the fourth quarter of 1939 and zero otherwise) . The coefficients of interest are β_k. Prior to second quarter of 1933 they capture the pre-treatment trends the lead (anticipatory) effects used to test if pre-treatment trends are parallel. After second quarter of 1933 they 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).

4.1 Results

The estimated coefficients β_k from the specification 1 are plotted in the figure 1. 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.

5 Synthetic Control Method

However, the parrarel trends assumption can sometiemes be violated. These issues can be adressed by sythetic control method (Abadie and Gardeazabal, 2003; Abadie et al.,

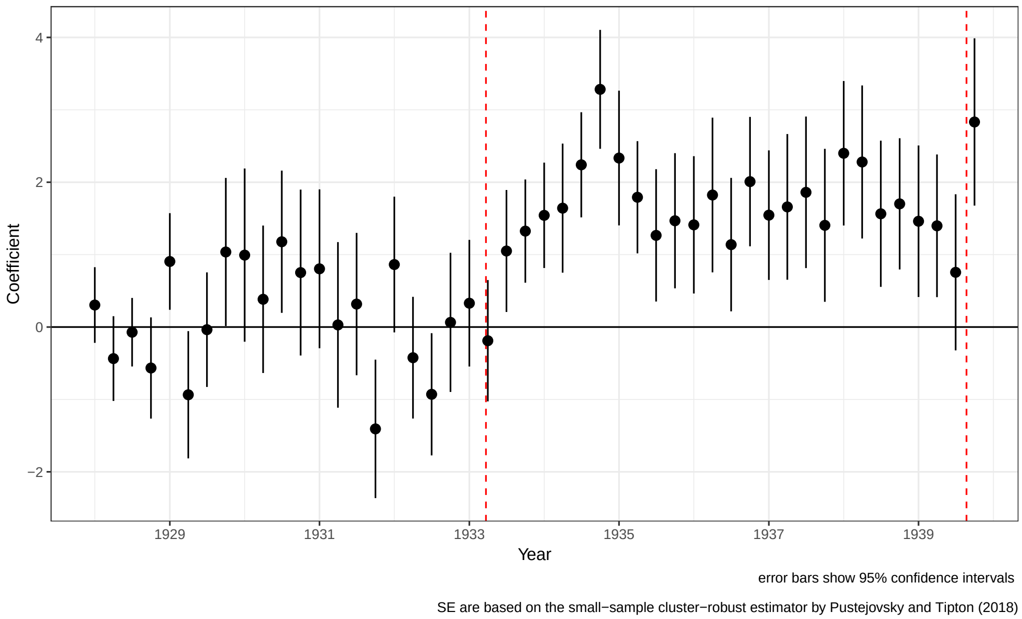


Figure 1: Estimates of coefficients β_k on the interactions of $\text{German}_i \cdot \text{YearQuart}_t^k$

2010)

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

$$Y_{1t} = Y_{1t}^N + \alpha_{1t} D_{1t} \tag{2}$$

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

$$Y_{1t}^N = \delta_t + \boldsymbol{\theta}_t \boldsymbol{Z}_i + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \epsilon_{it} \tag{3}$$

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-

Table 1: Synthetic German minority weights

Ethnic group	Weight
Ossetian	0.39
Tatar	0.23
Polish	0.14
Greek	0.09
Kabardin	0.07
Chechen	0.03
Lithuanian	0.02
Ukrainian	0.01

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.:

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

5.1 Results

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 1 (ethnic groups with zero weight are not shown). The highest contribution in the synthetic German minority have the 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 2 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 3a) 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.

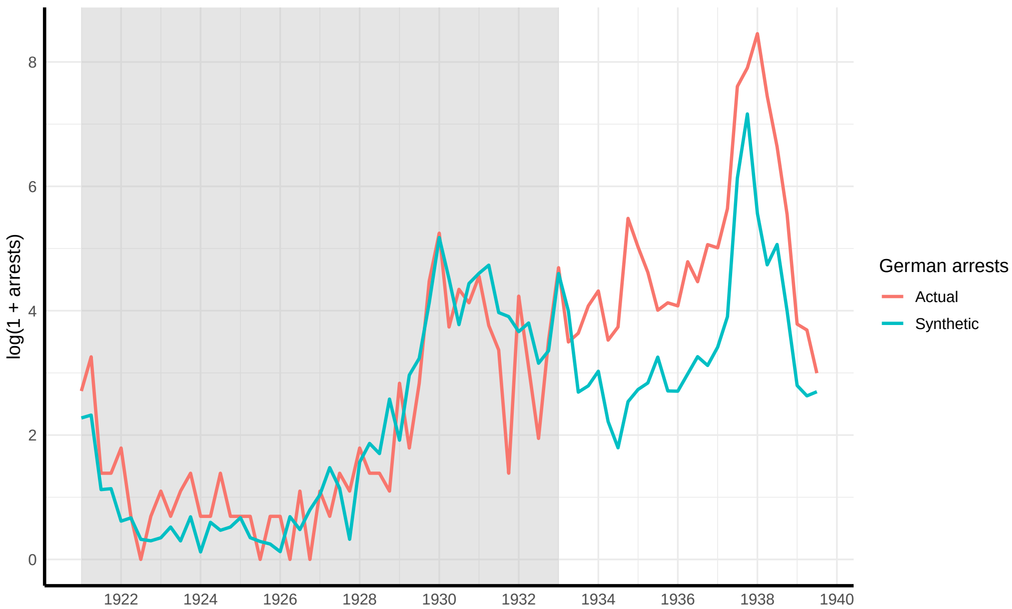
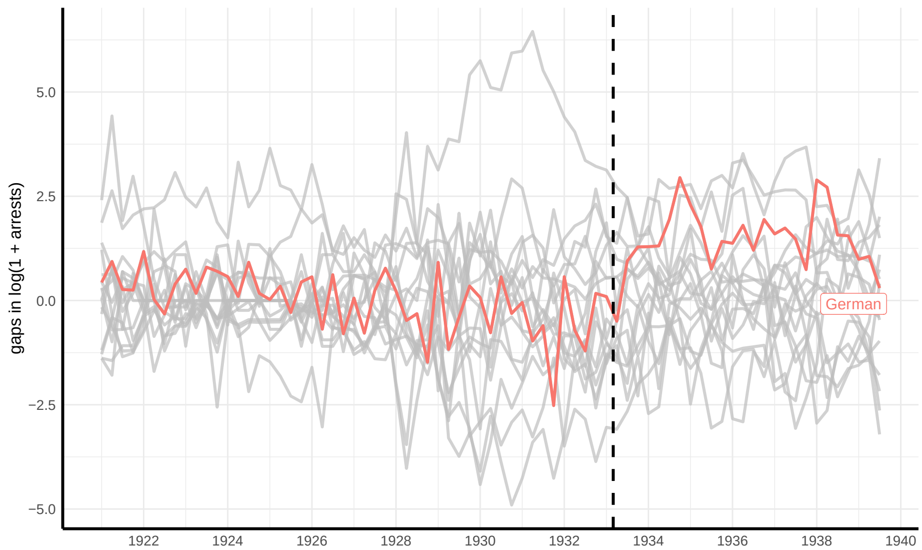


Figure 2: Comparison plot

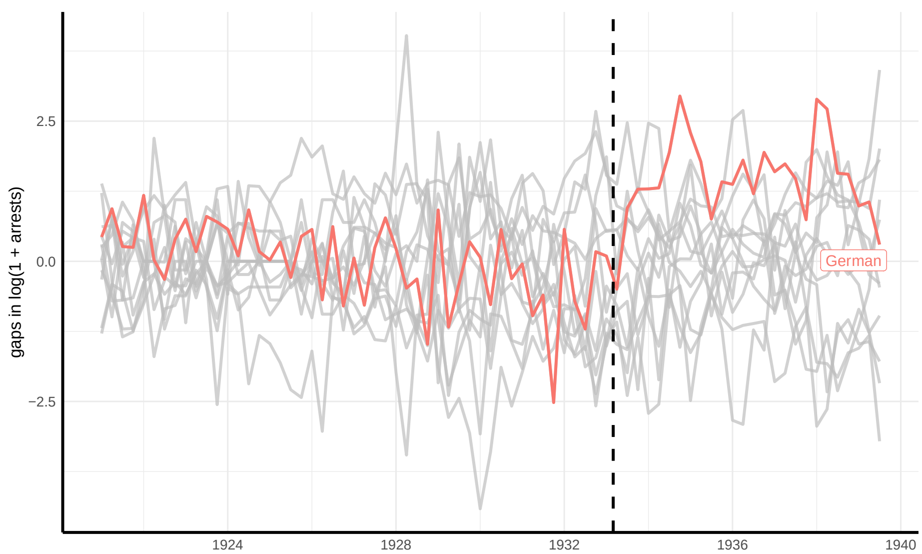
The synthetic control method, however, by itself does not provide us with any measure of uncertainty of significance. This has been addressed by performing placebo tests (Abadie et al., 2010). Synthetic control method is applied iteratively to every ethnicity in the donor pool as if they were treated. 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 3a 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 (Tatars, ...). The resulting plot is shown in the figure 3b. The post-1933 gaps in German arrests now stand out more clearly.



(a) All ethnic groups



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

Figure 3: Gaps between synthetic control and actual values for placebo tests

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 4. 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).

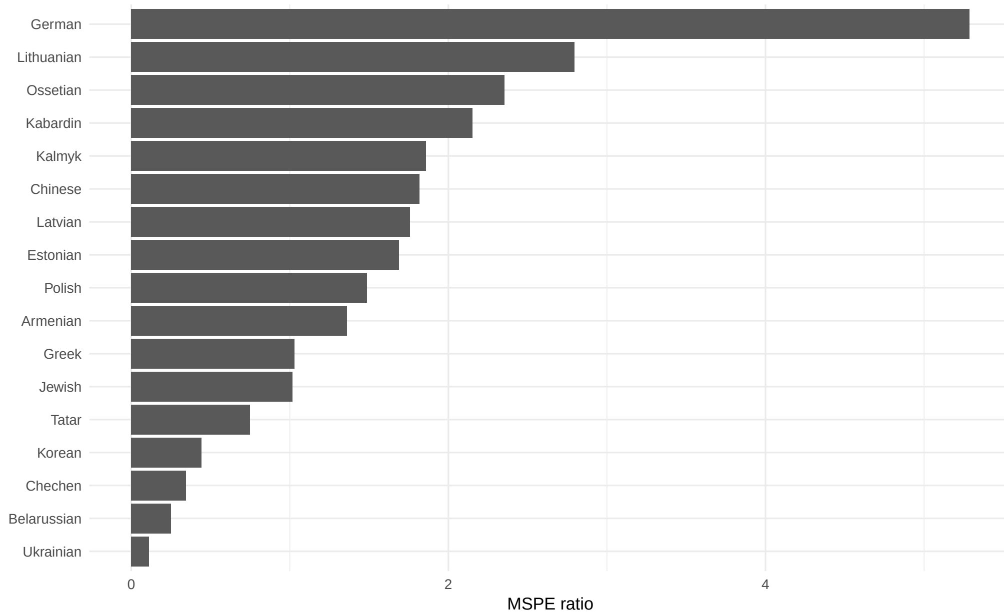


Figure 4: Ratios of post-treatment MSPE to pre-treatment MSPE

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.

Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country, *American Economic Review* **93**(1): 113–132.
URL: *https://www.aeaweb.org/articles?id=10.1257/000282803321455188*

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

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.

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 the domestic democratic peace*, Cambridge University Press.

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

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*

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*

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.

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.

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. (n.d.). Demographic Engineering and International Conflict: Evidence from China and the Former USSR, *International Organization* .

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.

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*

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.

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

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

Table 2: Total arrest by ethnicity, 1921-1958

Ethnicity	Number of arrests
Armenian	2228
Belarussian	66226
Estonian	7508
German	37812
Greek	1508
Chechen	723
Chinese	6317
Jewish	28900
Kabardin	2162
Kalmyk	5405
Korean	2482
Latvian	13208
Lithuanian	2666
Ossetian	2812
Polish	54022
Tatar	26294
Ukrainian	49306

Table 3: Naive Bayes Performance Measures by Ethnicity

Ethnicity	Sensitivity	Specificity
Armenian	0.573	0.998
Belarussian	0.430	0.975
Estonian	0.576	0.992
German	0.773	0.981
Greek	0.572	0.986
Chechen	0.682	0.962
Chinese	0.817	0.999
Jewish	0.652	0.995
Kabardin	0.741	0.997
Kalmyk	0.608	0.995
Korean	0.787	0.998
Latvian	0.546	0.993
Lithuanian	0.385	0.992
Ossetian	0.722	0.994
Polish	0.647	0.978
Russian	0.873	0.853
Tatar	0.725	1.000
Ukrainian	0.360	0.975

Table 4: Confusion Matrix (based on 10-fold cross-validation) - Counts

	Reference																
Prediction	Armenian	Belarussian	Estonian	German	Greek	Chechen	Chinese	Jewish	Kabardin	Kalmyk	Korean	Latvian	Lithuanian	Ossetian	Polish	Russian	Tatar	Ukrainian
Armenian	1341	275	16	73	7	0	3	193	9	46	2	15	13	5	194	810	17	233
Belarussian	27	29240	66	269	30	8	5	460	10	152	11	290	175	23	7240	9378	37	3847
Estonian	37	538	4379	1391	27	3	6	477	9	97	8	1056	79	38	990	2422	41	411
German	43	1113	759	35325	22	12	13	1654	59	452	14	1281	108	31	2030	8165	28	1069
Greek	78	1495	136	528	887	3	12	507	10	292	18	272	58	65	784	7373	55	1656
Chechen	288	1507	357	1879	67	477	600	4799	940	4588	104	666	256	239	4955	8399	5269	1491
Chinese	2	26	1	28	0	20	6358	20	2	63	676	8	4	1	38	169	22	34
Jewish	54	513	88	599	10	5	12	25795	9	80	2	121	24	23	454	2068	45	542
Kabardin	15	161	19	192	7	20	5	181	3827	166	1	29	7	61	119	1214	195	130
Kalmyk	13	328	12	200	3	6	16	134	21	19739	1	37	16	13	251	3239	17	327
Korean	5	102	17	73	5	0	414	60	3	71	3594	13	2	4	106	660	13	109
Latvian	15	509	658	1211	8	5	6	259	11	102	11	7311	98	14	1434	1926	9	482
Lithuanian	28	788	128	401	30	3	25	381	3	508	14	268	1223	22	2238	2003	181	552
Ossetian	21	430	55	201	18	16	42	806	137	277	8	95	39	2099	586	2147	306	373
Polish	49	8204	133	1026	21	6	9	1249	25	80	8	630	751	17	51763	5290	45	1972
Russian	258	18804	701	1964	347	83	225	2038	57	5446	86	1128	288	205	5058	483478	1095	20345
Tatar	1	1	0	2	1	27	19	8	27	26	0	1	0	2	0	107	19591	3
Ukrainian	64	3938	72	345	62	5	10	549	8	275	11	180	37	44	1713	15162	40	18924

Table 5: Confusion Matrix (based on 10-fold cross-validation) - Proportions

	Reference																	
Prediction	Armenian	Belarussian	Estonian	German	Greek	Chechen	Chinese	Jewish	Kabardin	Kalmyk	Korean	Latvian	Lithuanian	Ossetian	Polish	Russian	Tatar	Ukrainian
Armenian	0.573	0.004	0.002	0.002	0.005	0.000	0.000	0.005	0.002	0.001	0.000	0.001	0.004	0.002	0.002	0.001	0.001	0.004
Belarussian	0.012	0.430	0.009	0.006	0.019	0.011	0.001	0.012	0.002	0.005	0.002	0.022	0.055	0.008	0.091	0.017	0.001	0.073
Estonian	0.016	0.008	0.576	0.030	0.017	0.004	0.001	0.012	0.002	0.003	0.002	0.079	0.025	0.013	0.012	0.004	0.002	0.008
German	0.018	0.016	0.100	0.773	0.014	0.017	0.002	0.042	0.011	0.014	0.003	0.096	0.034	0.011	0.025	0.015	0.001	0.020
Greek	0.033	0.022	0.018	0.012	0.572	0.004	0.002	0.013	0.002	0.009	0.004	0.020	0.018	0.022	0.010	0.013	0.002	0.032
Chechen	0.123	0.022	0.047	0.041	0.043	0.682	0.077	0.121	0.182	0.141	0.023	0.050	0.081	0.082	0.062	0.015	0.195	0.028
Chinese	0.001	0.000	0.000	0.001	0.000	0.029	0.817	0.001	0.000	0.002	0.148	0.001	0.001	0.000	0.000	0.000	0.001	0.001
Jewish	0.023	0.008	0.012	0.013	0.006	0.007	0.002	0.652	0.002	0.002	0.000	0.009	0.008	0.008	0.006	0.004	0.002	0.010
Kabardin	0.006	0.002	0.003	0.004	0.005	0.029	0.001	0.005	0.741	0.005	0.000	0.002	0.002	0.021	0.001	0.002	0.007	0.002
Kalmyk	0.006	0.005	0.002	0.004	0.002	0.009	0.002	0.003	0.004	0.608	0.000	0.003	0.005	0.004	0.003	0.006	0.001	0.006
Korean	0.002	0.002	0.002	0.002	0.003	0.000	0.053	0.002	0.001	0.002	0.787	0.001	0.001	0.001	0.001	0.001	0.000	0.002
Latvian	0.006	0.007	0.087	0.026	0.005	0.007	0.001	0.007	0.002	0.003	0.002	0.546	0.031	0.005	0.018	0.003	0.000	0.009
Lithuanian	0.012	0.012	0.017	0.009	0.019	0.004	0.003	0.010	0.001	0.016	0.003	0.020	0.385	0.008	0.028	0.004	0.007	0.011
Ossetian	0.009	0.006	0.007	0.004	0.012	0.023	0.005	0.020	0.027	0.009	0.002	0.007	0.012	0.722	0.007	0.004	0.011	0.007
Polish	0.021	0.121	0.018	0.022	0.014	0.009	0.001	0.032	0.005	0.002	0.002	0.047	0.236	0.006	0.647	0.010	0.002	0.038
Russian	0.110	0.277	0.092	0.043	0.224	0.119	0.029	0.052	0.011	0.168	0.019	0.084	0.091	0.071	0.063	0.873	0.041	0.388
Tatar	0.000	0.000	0.000	0.000	0.001	0.039	0.002	0.000	0.005	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.725	0.000
Ukrainian	0.027	0.058	0.009	0.008	0.040	0.007	0.001	0.014	0.002	0.008	0.002	0.013	0.012	0.015	0.021	0.027	0.001	0.360

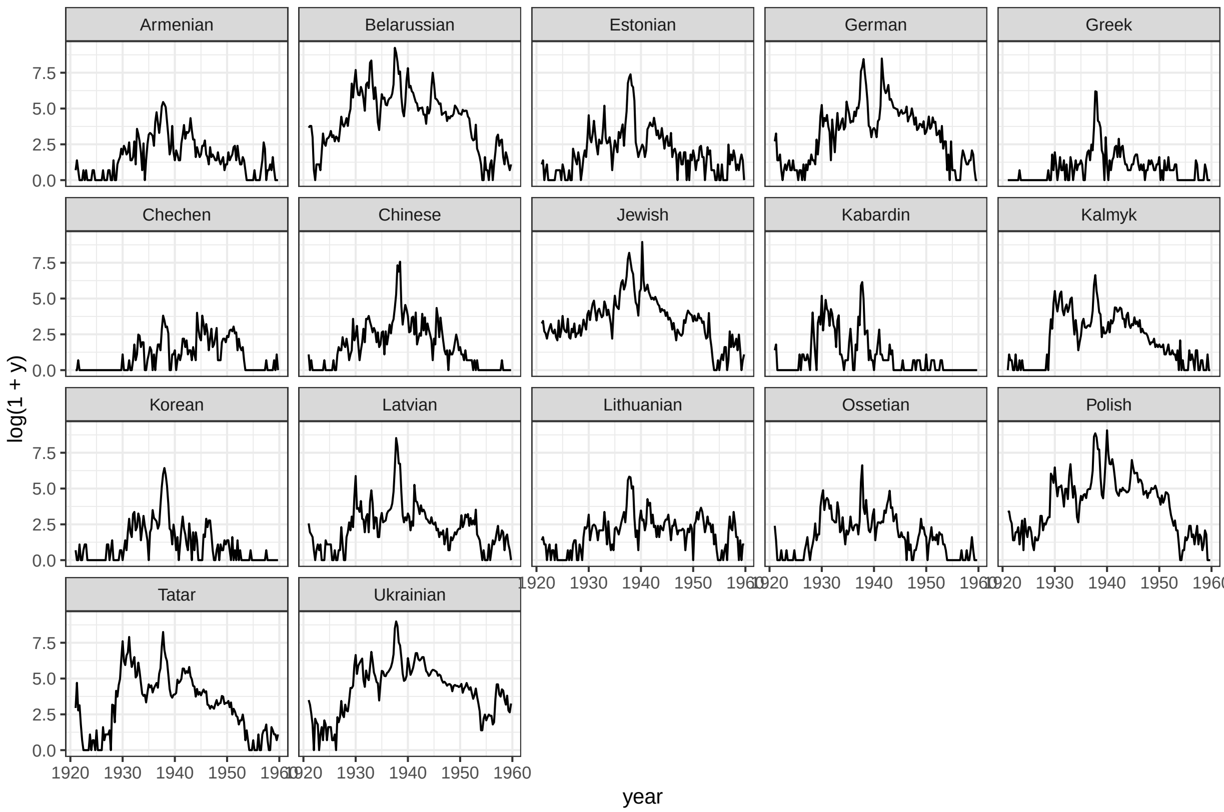


Figure 5: Arrests by ethnicity and year (in $log(1 + \text{arrests}_{it})$)

Table 6: Pre-treatment characteristics of ethnic groups in the USSR for SC

Ethnic group	Total population	Linguistic similarity to Russian	Urbanization rate
Armenian	1 567 568	1	35.45
Belarussian	4 738 923	4	10.32
Estonian	154 666	0	23.00
German	1 238 549	1	14.92
Greek	213 765	1	21.21
Chechen	318 522	0	0.98
Chinese	10 247	0	64.87
Jewish	2 599 973	1	82.43
Kabardin	139 925	0	1.27
Kalmyk	129 321	0	1.29
Korean	86 999	0	10.52
Latvian	141 703	2	42.31
Lithuanian	41 463	2	63.16
Ossetian	272 272	1	7.86
Polish	782 334	3	32.75
Tatar	2 916 536	0	15.48
Ukrainian	31 194 976	4	10.54

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).