NBER WORKING PAPER SERIES

THE CAUSES OF UKRAINIAN FAMINE MORTALITY, 1932-33

Andrei Markevich
Natalya Naumenko
Nancy Qian

Working Paper 29089
http://www.nber.org/papers/w29089

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2021, Revised November 2022

Previously circulated as "The Political-Economic Causes of the Soviet Great Famine, 1932–33." We thank Daron Acemoglu, Maxim Ananyev, Michael Ellman, Ruben Enikolopov, Jeff Frieden, Scott Gehlbach, Mikhail Golosov, Sergei Guriev, Mark Harrison, Oleg Khlevniuk, Alexey Makarin, Victor Malein, Joel Mokyr, Cormac O'Grada, Gerard Padro-i-Miquel, Steven Nafziger, Maria Petrova, Sakari Saaritsa, Torsten Santavirta, Marco Tabellini, Chris Udry and Ekaterina Zhuravskaya for their insights; the participants of the Harvard/MIT Development Workshop, NYU Development Workshop, LSE Sticerd AppliedWorkshop, Northwestern Economic HistoryWorkshop, Fudan University SOE AppliedWorkshop, Trinity College Applied EconomicsWorkshop, CERGE-EI Applied EconomicsWorkshop, EIEFWorkshop, HECER (Helsinki Center of Economic Research) Labor and Public Economics seminar, Helsinki economic and social history seminar, Northwestern Economic History Lunch, Northwestern Development Lunch, George Mason Economic History Lunch, the NBER Political Economy Workshop, the Ridge Conference, the ASSA/AEA On the Role of and Culture and Community 2020 session, 2020 Summer Workshop in the Economic History and Historical Political Economy of Russia, Kiel/CEPR Conference on Geopolitics. For excellent research assistance, we thank Pavel Bacherikov, Nikita Karpov, Ivan Korolev, Sergio Lopez-Araiza Bejar, Kaman Lyu, Carlo Medici, Ludovica Mosillo. All mistakes are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Causes of Ukrainian Famine Mortality, 1932-33
Andrei Markevich, Natalya Naumenko, and Nancy Qian
NBER Working Paper No. 29089
July 2021, Revised November 2022
JEL No. N14,O1,O13,P16

ABSTRACT

We construct a large dataset to understand the causes of high Ukrainian mortality during the Great Soviet famine (1932-33). We document that holding per capita grain production, urbanization, and other factors constant, famine mortality rate was increasing in pre-famine ethnic Ukrainian population share across regions, even outside of Ukraine. Government grain procurement as a share of production is also increasing in pre-famine ethnic Ukrainian population share. These and other results imply that Ukrainian bias in Soviet policy was the main contributor to high Ukrainian famine mortality, and rule out alternative explanations such as bad weather or other exogenous factors.

Andrei Markevich
New Economics School
45 Skolkovskoe shosse
Moscow
Russia
ammarkevich@gmail.com

Natalya Naumenko
Department of Economics
George Mason University
Carow Hall, Office 10
4500 Roberts Rd
Faifax, VA 22032
United States
nata.naumenko@gmail.com

Nancy Qian
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, Il 60208
and NBER
nancy.qian@kellogg.northwestern.edu

1 Introduction

During the Great Soviet Famine (1932-33), the second worst famine in the 20th Century in terms of total deaths, approximately seven million people perished.[1] Ukrainians call the famine *Holodomor* ("to die by starvation") because of the disproportionally high mortality rates they suffered. Ethnic Ukrainians, who were one-fifth of Soviet population, contributed to approximately forty percent of these deaths. 7.5% to 11.3% of the ethnic Ukrainian peasantry perished from famine. Famine mortality in Ukraine, where most ethnic Ukrainians lived, was four to six times higher than in Russia.[2] The causes of Ukrainian famine mortality have been a subject of intense controversy and academic debate.

On the one side is the *Ukrainian bias* view, which argues that higher Ukrainian famine mortality was an outcome of Soviet bias against Ukrainians. Conquest (1986) famously claimed that the famine was a "terror" intentionally waged by the Soviet government on Ukrainians. This is motivated by the following facts. Food availability in the centrally planned economy was determined by government policy, which procured food from rural areas to feed urban workers, export or store in reserves. The main Soviet agricultural policy of collectivizing agricultural production was highly unpopular amongst the peasantry. The Bolsheviks were especially wary of Ukrainian resistance because Ukrainians were the largest ethnic group in agriculturally productive regions, had a strong group identity and a history of resisting the Bolsheviks. Thus, the Bolsheviks repressed Ukrainians in order to control agricultural production, which constituted almost half of Soviet GDP (e.g., Graziosi, 2015).[3] The central claims of the Ukrainian bias view are that regional food production was sufficient for sustaining the Ukrainian population and that the famine occurred because the state procured too much food from Ukrainians.

In direct opposition, the *no bias* view argues that there was no Ukrainian bias in Soviet policies and high Ukrainian famine mortality was an outcome of circumstance and bad luck. This view draws on the facts that, in an intentional departure from the Tsarist regime, Communist ideology theoretically held all ethnicities to be equal; and that historians have not recovered a "smoking gun" document that Stalin "ordered" a famine. It is supported by the observation that severe famine also occurred in agriculturally productive regions outside of Ukraine (e.g., Kondrashin, 2008). Davies and Wheatcroft (2009) document that the production of grain, the most important staple crop in the Soviet Union, fell during the famine.[4] According to the no bias view, high Ukrainian famine mortality was due to low agricultural production and exogenous factors that influenced production, such as bad weather; all else equal, there was no bias against Ukrainians.

This debate has been at an *impasse* due to the lack of empirical evidence. Our study aims to address this. We ask one question: did Soviet bias against Ukrainians contribute to high Ukrainian famine mortality? To answer this question, we examine the relationship between famine mortality and the share of pre-famine ethnic Ukrainians. To isolate the contribution of Ukrainian bias to famine mortality, we control

[1] The Chinese Great Famine (1959-61) experienced a similar mortality rate and a higher number of total deaths. The Background Section discusses the estimated range in Soviet famine deaths.

[2] Note that approximately 1 to 1.5 million famine deaths occurred in Kazakhstan, who were mostly ethnic Kazakhs. We do not study Kazakh mortality because there are no reliable mortality data from Kazakhstan during the famine era. See the Background for a discussion of famine mortality estimates.

[3] Conquest (1986), Ellman (2007), and Mace (2004) offer variations of this argument. The discussion here focuses on the scholarly debate on the causes of Holodomor. In the Conclusion, we also discuss the debate amongst policymakers and the public.

[4] Also, see Kotkin (2017).

for regional food production and many other characteristics. To provide positive evidence of Ukrainian bias in Soviet policy, we examine the relationship between the share of pre-famine ethnic Ukrainians and the most important policy for determining food availability: centrally planned grain procurement.

The main empirical challenge for our study is the lack of systematic and disaggregated data, which are needed to evaluate competing hypotheses. We address this by constructing the largest and most comprehensive dataset for inter-World War Soviet Union, from 1922 to 1940. We construct panel data that include the following information: mortality, natality, ethnic composition, urbanization, weather, administrative capacity, realized grain production and procurement, planned targets for production and procurement, and historical institutions. We draw mainly from archival sources made available after the fall of the USSR. The main sample we use for the regression analysis includes the three largest and most populous Soviet republics: Belarus, Russia, and Ukraine. Our main analysis uses a province-level panel, for which we have a long time horizon and a large set of variables. We supplement this with a more granular district-level panel, for which we have fewer years and variables.

The paper proceeds as follows. First, we motivate the main analysis with a simple republic-level food accounting exercise. Following existing studies of famine, we focus on grain production. We show that 1932 grain production in Ukraine was lower than in earlier years, but it was still enough to avoid severe famine if the regime had not procured grain from Ukraine; moreover, aggregate production in the rest of the Soviet Union was sufficient for sustaining the population without transfers from Ukraine.[5] Thus, low regional production and the contributors to production cannot explain severe famine in Ukraine; the main cause is the over-procurement of food from Ukraine.

Second, the main empirical analysis tests the two opposing views on the causes of high Ukrainian famine mortality with disaggregated region-level panel data. The *Ukrainian bias* view predicts that for two regions which experienced the same weather and produced the same amount of food, famine mortality is higher in the region with more ethnic Ukrainians. In contrast, the *no bias* view predicts that, other things being equal, famine mortality is similar in the two regions. The baseline estimate regresses mortality for a given province and year on the pre-famine ethnic Ukrainian population share interacted with a famine-year dummy variable and additional controls. We infer Ukrainian-specific famine mortality from this relationship since there are no ethnic-specific mortality data. We control for grain production and its interaction with the famine dummy variable so that we compare provinces with similar amounts of food production, and allow the influence of production to differ flexibly for the famine year. We control for urban population share and its interaction with the famine dummy variable to account for urban-rural differences in access to food and other factors that affect mortality. Province fixed effects control for time-invariant differences across regions (e.g., mortality is always higher in some regions). Year fixed effects control for changes over time that affect all provinces similarly (e.g., general improvements in health care).

We find that the interaction of ethnic Ukrainian population share and the famine dummy variable is positive and large in magnitude. This supports the Ukrainian bias view and rejects the no bias view.

There are several potential issues to consider for interpreting our findings. One concern is that official

[5]In other words, this means that if production was equally distributed across the population, production in the other Soviet republics would have been enough for sustaining their population.

figures of grain production from the early 1930s are exaggerated. This is a problem for our preferred interpretation if the exaggeration was more pronounced in areas with higher ethnic Ukrainian population share. In the accounting exercise, we address this by using previously classified data sources, which are only available for a few years, to revise official production estimates (see Section 3). To maximize the time span of the mortality analysis, we address this problem another way: in the regressions, we control for production predicted by exogenous factors such as weather and geography instead of realized production. We use pre-Soviet data to estimate a production function, and then use the production function and Soviet era parameter values to predict grain production. Our results are similar if we alternatively control for the corrected grain production estimates used for the accounting exercise (but the sample size is much smaller) or if we directly control for weather instead of the predicted grain. The main challenge in controlling for predicted production is that policies prior to the famine that lowered agricultural productivity may have been more intensely implemented in Ukrainian areas. This could cause predicted production to overstate production in Ukrainian areas. Two policies fit this description: *dekulakization*, which aimed to eliminate wealthy peasants who had resisted Soviet agricultural reforms; and the drop in livestock, which occurred a few years prior to the famine in response to the collectivization of productive assets. Lowering the supply of skilled farmers and livestock may have lowered agricultural productivity. We address this issue by showing that our main results are robust to controlling for the interaction of the famine dummy variable and different measures of *dekulakization* and the drop in livestock.

Another concern is measurement error in mortality, our main measure of famine severity, and ethnic Ukrainian population share, which we take from the 1926 population census. We demonstrate that our estimates are qualitatively similar when using a large number of alternative measures of famine severity, such as birth rates, birth cohort size measured in the 1939 population census; and alternative measures of ethnic Ukrainian population share, such as ethnicity and mother tongue reported in the 1926 and 1897 population censuses.

The third concern for interpreting the main result is omitted variables: pre-famine ethnic Ukrainian population share may be correlated with other factors that contribute to famine mortality. We provide a large body of evidence against this. We show that our main finding is robust to controlling for a large number of additional variables, such as demographic structure, suitability for the cultivation of other crops (e.g., potatoes), political loyalty to the regime, administrative capacity and historical economic and institutional variables. We also use the 1892 famine, the last large famine under the Tsarist regime, as an alternative experiment. This is important given the recent evidence that slow-moving features of a society, such as cultural or social norms, historically affect famine intensity in Europe (Buggle and Durante, 2021). In that case, Ukrainian bias in 1932-33 famine mortality will reflect these other factors instead of Ukrainian bias in Soviet policy. We find that ethnic Ukrainian population share was uncorrelated with 1892 famine mortality. Thus, the Ukrainian bias revealed by the baseline estimates is unique to the Soviet famine of 1932-33 and unlikely to be driven by slow-moving unobserved differences between Ukrainians and other ethnic groups.

Another concern is that our estimates are driven by outliers or a few influential observations. This stems from the fact that there are few provinces and ethnic Ukrainian population is concentrated. To address this concern, we show that the results are qualitatively similar if we omit Ukraine and other provinces where

most ethnic Ukrainians are concentrated. We also conduct random permutation tests to show that our results are highly unlikely to be driven by coincidence.

Finally, one may be concerned that famine mortality in regions with high ethnic Ukrainian shares are driven by the mortality of ethnic Russians, who make up most of the non-Ukrainians in our sample (the ecological fallacy). There are no known anecdotal accounts of such mortality patterns. Moreover, the results for ethnic Ukrainian population share after the famine that we discuss later go against this concern.

The main province-level results support the Ukrainian bias view and reject the no bias view. The district-level panel data allow us to strengthen this interpretation in two ways. Since Soviet economic policies were centrally planned and implemented top-down through the bureaucratic hierarchy, central planners typically used similar formulas to assign production targets across provinces and across smaller administrative units (districts) within a province. Thus, if the province-level relationship between famine mortality and ethnic Ukrainian population share reflects systematic bias in Soviet policy, we should find a similar relationship across districts. We estimate the baseline equation with the district-level data and the additional control of province-year fixed effects, which control for all time-varying province-level differences. We find that there is Ukrainian bias in famine mortality across districts within provinces.

The district-level data also allow us to investigate whether the bias against Ukrainians was delineated along ethnic or administrative lines. Given the lack of ethnic-specific mortality data, existing discussions about Ukrainian bias usually compare mortality in Ukraine versus mortality in Russia and other republics. This ignores the fact that many ethnic Ukrainians lived in other republics of the USSR. We overcome this problem by inferring ethnic Ukrainian famine mortality from the relationship between famine mortality and pre-famine ethnic Ukrainian population share. In the main province-level analysis, we find qualitatively similar estimates when we exclude Ukraine. This means that there was systematic bias against ethnic Ukrainians in famine mortality across the Soviet Union, even outside Ukraine. The district-level data allow us to take this investigation a step further by investigating whether there is a border effect for famine mortality along the Ukrainian-Russian border. Consistent with anecdotal accounts, we document a discrete decline in famine mortality rates when crossing the border from Ukraine to Russia. However, this decline disappears when we control for the ethnic Ukrainian population share in each district. These results are interesting for several reasons. They emphasize that the Ukrainian bias was delineated along ethnic and not administrative lines, and was systematic across the Soviet Union. They illustrate how observations can be misleading without controlling for confounding factors. The presence of a strong border effect when not controlling for ethnic Ukrainian population share is consistent with conventional wisdom that the movement of food and population between famine stricken regions and regions that were relatively better off was limited. Henceforth, we return to using the province-level data.

A back-of-the-envelope calculation implies that overall bias against ethnic Ukrainians explains up to 92% of total famine mortality in Ukraine and 77% in the three republics in our sample.[6]

We supplement the main analysis with several additional results. The first is to examine whether there is Ukrainian bias in the most important state policy for determining food availability: centrally planned grain

[6]These estimates should not be interpreted literally, but as illustrative of the importance of policy bias against ethnic Ukrainians in explaining famine mortality. See the paper for more details.

procurement. Positive statistical evidence on the presence of such bias in this policy can help address the lack of smoking-gun documents and the fact that the documentary evidence is only partial. We estimate a similar equation as the mortality baseline. We find that all else equal, grain procurement as a share of production is higher in regions with higher ethnic Ukrainian population shares prior to the famine. In other words, for two places that produced similar amounts of food, more was procured from the place with a higher share of ethnic Ukrainians.[7] Analogously, we find that food retention, which is production minus procurement, is decreasing with ethnic Ukrainian population share. We also document that higher retention is associated with lower mortality. Taken together, these estimates provide positive evidence that Ukrainian bias in the main government policy for determining food availability contributed to higher Ukrainian famine mortality. We also collect production and procurement targets from the First Five Year Plan published in 1928 and document that, conditional on production targets, procurement and retention targets were biased against ethnic Ukrainians. Thus, Ukrainian bias in grain procurement was planned.[8] A simple quantification exercise shows that Ukrainian bias in centrally planned procurement explains approximately half of the total effect of Ukrainian bias on famine mortality. This is consistent with the importance of centrally planned grain procurement in determining food availability, as well as the presence of Ukrainian bias in other policies that affect famine mortality.[9]

This study focuses on the role of Ukrainian bias as a cause of high Ukrainian famine mortality. For a sense of the long-run impact of Ukrainian bias in the famine, the last empirical exercise uses population censuses from 1897 to 2002 to investigate the long-run demographic consequences. We document that in provinces with high pre-famine Ukrainian population share that suffered higher famine mortality, total population had broadly recovered by the 1950s and 1960s, but the size and share of ethnic Ukrainian population was permanently lower after the famine.

Our study is the first to provide rigorous empirical evidence that systematic bias against ethnic Ukrainians in Soviet Policy was a key contributor to high Ukrainian famine mortality during 1932-33. An important question for future research is to better understand the regime's motivation for repressing ethnic Ukrainians. To shed some light on this question, we provide a speculative discussion at the end of the paper and present suggestive evidence consistent with the political economic explanation offered by historians: the Bolsheviks repressed Ukrainians to control agriculture. There are other possible explanations. It is beyond the scope of this paper to be conclusive on this point.

This study contributes to several literatures. First, it adds to the literature on ethnic conflict (e.g., Alesina and La Ferrara, 2005; Esteban et al., 2012; Montalvo and Reynal-Querol, 2005). Recent studies of Eastern Europe and the former Soviet Union provide important evidence on the impact of group conflicts in contexts such as the abolition of serfdom (Buggle and Nafziger, 2019; Markevich and Zhuravskaya, 2018), forced

[7] For these estimates, we use the corrected production figures from the accounting exercise in Section 3.

[8] Ukrainian bias in planned targets does not necessarily mean that the famine was planned because the production targets published in 1928 were unrealistically high and reductions in consumption from high base levels do not result in mortality. The results are consistent with both: *i*) a policy that exposed ethnic Ukrainians to greater famine risk by leaving them with less food, but with no *ex ante* intention of reducing the ethnic Ukrainian population size; and *ii*) a policy that intended to reduce ethnic Ukrainian population size by killing ethnic Ukrainians. We discuss this in detail in the paper.

[9] For example, there may be Ukrainian bias in the disbursement of aid or the degree to which migration out of famine-stricken areas was allowed. Also, in addition to grain, the Soviet government procured other agricultural products, about which we do not have data. See the paper for a more detailed discussion.

migration (Bauer et al., 2013; Becker et al., 2020), peasant rebellions (Castañeda Dower et al., 2018; Finkel and Gehlbach, 2020), mass repressions (Talibova and Zhukov, 2018), anti-Semitism (Grosfeld et al., 2013; Acemoglu et al., 2019) and ethnic diversity as a legacy of historical ethnic tensions (Egorov et al., 2020). There are fewer studies about the causes of ethnic conflict. Grosfeld et al. (2020) documents the economic causes of anti-Jewish pogroms during 1800-1927. DellaVigna et al. (2014) and Adena et al. (2015) find that exposure to propaganda heightens nationalist sentiments along the Croatian-Serbian border and deepens ethnic divides in Nazi Germany.[10]

We also add to studies that evaluate the successes and failures of early Soviet economic policies (Cheremukhin et al., 2017) and the long-run effects of Communism (Gorodnichenko and Roland, 2017; Roland, 2010).[11] Our findings complement recent empirical evidence that Ukrainian-Russian ethnic delineations affect firm behavior after the 2014 Russian annexation of Crimea (Korovkin and Makarin, forthcoming), and that historical exposure to the famine affects political preferences of Ukrainians in the post-Soviet era (Rozenas and Zhukov, 2019).

Finally, our findings contribute to studies of the causes of famine in the post-Industrial era. Sen (1981) famously argues that the central cause of 20th century famines is the unequal distribution of food from political elites to those who lack entitlement, and not low aggregate production. Our results strongly support this thesis by providing evidence on the political-economic process that led to the second largest famine in the 20th century. We complement recent empirical analyses of China (e.g., Li and Yang, 2005; Meng et al., 2015), India (e.g., Sen, 1981; Burgess and Donaldson, 2017) and Ukraine (Naumenko, 2021).[12] These earlier works do not study ethnic bias.

This paper is organized as follows. Section 2 summarizes the historical background. Section 3 presents the food accounting exercise. Section 4 presents the main mortality analysis. Section 5 presents additional results. Section 6 concludes.

2 Background

2.1 Soviet National [Ethnic] Policy

In a departure from explicit ethnic discrimination during the Tsarist regime, official Bolshevik ideology held all ethnicities within the Soviet Union to be equal (Martin, 2001). In practice, the Soviet regime was Russo-centric: the capital was in Russia and all republics used Russian as one of the main languages. From the beginning of the Soviet rule in 1917, the regime was wary of nationalist (ethnically delineated) sentiments that threatened to undermine its existence. The Civil War (1918-1920) revealed the support for separatist nationalist movements. The power of national groups was partly a result of the historical concentration and segregation of ethnic groups in the former Russian empire, which continued after the establishment of the

[10] There are numerous studies about the role of media on ethnic conflict in other contexts. See Petrova and Yanagizawa-Drott (2016) for a review.

[11] Cheremukhin et al. (2017) provides macro-calibration evidence on the effect of early Soviet Industrial policy, but excludes the cost of famine because of data limitations. See Zhuravskaya et al. (forthcoming) for an overview of studies of Russian economic history.

[12] See Ó Gráda (2009) for a discussion of major famines in history. Naumenko (2021) documents a positive association between collectivization and 1933 famine mortality in a cross-section of districts in Ukraine.

USSR, especially in rural areas. Villages and larger geographical units were usually dominated by one large majority group and made it easier to coordinate within ethnic groups. A region (e.g., province) with a large share of ethnic Ukrainians is usually one that contains a large number of sub-units (e.g., villages) with ethnic Ukrainian majorities.

In the 1926 Census, 23.2 million ethnic Ukrainians lived in Ukraine and an additional 7.9 million lived in Russia and Belarus. Ukrainians constituted 21% of total Soviet population and made up the second largest ethnic group. Russians, the largest group, constituted 53% of the population. However, in regions that produced large food surpluses that the government designated as "grain-surplus" areas, Ukrainians were the largest group (43.8%) and Russians were a close second (41.9%). 89% of ethnic Ukrainians lived in rural areas.

To appease nationalists, the Bolsheviks launched a policy of indigenization (*korenizatsiya*) in 1923. In regions where minorities constituted the local majority, it encouraged schooling and the publication of books in native languages such as Ukrainian, promoted native culture (national literature, theaters, museums, etc.), required local government affairs to be implemented in the native language, and promoted locals into leadership positions in the government and industry. Government administration was also organized along ethnic/national lines. The Soviets established a hierarchy of national autonomous administrative units: republics — provinces — districts — village soviets. This system (inadvertently) deepened ethnic delineations, especially in rural areas that were already ethnically segregated.

There are several important implications for our study. First, ethnic identities were salient for the population and in state policy. Second, the regime could easily target specific ethnic groups by identifying geographic regions where they were concentrated even though Soviet peasants did not carry internal passports that included ethnic information until many years after the famine. For example, the 1926 population census that we use to measure pre-famine ethnic Ukrainian population share was collected by the government and available to central planners.

Third, the increasing salience of ethnicity was viewed as problematic for the regime even as it was conceived, especially in rural areas. As early as 1925, Stalin stated that "the national [ethnic] question [is], in essence, a peasant question" (Stalin, 15 April 1925 as quoted in Graziosi, 2015). According to Nikita Khrushchev, Lazar Kaganovich – the Stalin's deputy in the communist party in the early 1930s – "was fond of saying that every Ukrainian is potentially a nationalist" (Kuromiya, 2008). The concern intensified when peasants, particularly Ukrainian peasants, strongly resisted collectivization in the years just before the famine. Nearly half of all peasant uprisings against collectivization took place in Ukraine, and we will later show that resistance was positively associated with ethnic Ukrainian share across provinces (see Section 5). The Soviet regime *de facto* terminated indigenization policy in the western parts of the USSR as famine deaths began to mount in the autumn of 1932 (Graziosi, 2015; Martin, 2001).

2.2 Soviet Economic Policy and the 1932-33 Famine

Early Soviet economic policy focused on industrialization. Boosting grain production was critical for the success of industrialization and the political survival of the Bolshevik regime (1917-1991). Approximately half of Soviet GDP in 1928 comprised of agriculture, a larger part of which was grain production. The state

procured grain from the peasantry and used it to feed industrial urban populations and the small share of rural population engaged in non-agricultural production (e.g., forestry). Procured grains were also exported and stored in national reserves.

The goal of collectivizing agriculture was to strengthen the control of the state over grain production and distribution. Collectivization was a bundle of policies that included the removal of private property and organizing peasants into large collective farms that the government could control directly. The government banned the trading of food and instead procured grain directly from collective farms (and the remaining individual peasants). In theory, peasants were meant to be left with enough food so that they could live and work. Production targets were mainly based on the regime's assessment of production potential – e.g., past production, geography and rural labor force. Procurement targets, absent political objectives such as Ukrainian bias, were based on production targets and population subsistence needs. Forced collectivization began in late 1929. By the summer of 1932, the share of rural households living in collective farms exceeded 60% in the USSR and was almost 70% in Ukraine (Davies and Wheatcroft, 2009).

The First Five-Year Plan was published in 1928 and stated unrealistically ambitious targets for agricultural production and procurement. The untenableness of the production and procurement targets became apparent after the harvest of 1931, which was lower than planned and average production. News of starvation traveled to Moscow, but the government did not lower procurement. In areas where the peasants retained too little food, seed stock was consumed to make up for the deficit. The reduced seed stock and weakened labor force, in turn, contributed to lower production in 1932. National production in 1932 was only slightly lower than 1931, but some areas, such as Ukraine, experienced a larger decline.

The government initially resisted reducing procurement targets in 1932 despite news of falling harvests. In the summer of 1932, Stalin received multiple reports indicating the reluctance of Party leaders at all levels in Ukraine to facilitate the starvation of so many peasants.[13] Stalin remained committed to maximizing procurement even as he curtailed the initial targets out of necessity. To fulfill the remaining procurement quota, Stalin sent his closest deputies, Vyacheslav Molotov and Lazar Kaganovich, neither of whom were Ukrainian, to Ukraine and North Caucasus, the two key grain-producing regions where most ethnic Ukrainians lived.

The Stalin-led central leadership claimed that shortages and famine were outcomes of intentional peasant resistance aimed to undermine agricultural collectivization, and that the peasants should be penalized for their subversion (Danilov et al., eds, 1999-2006; Davies and Wheatcroft, 2009).[14] On December 14, 1932, the Politburo of the Communist Party and the Soviet government issued a classified decree in which the Bolshevik leaders accused Ukrainian nationalists within the Communist Party and local bureaucracy of sabotaging grain procurement. The decree required regional authorities in Ukraine (as well as in the North Caucasus and the Western region) to "crush" any resistance of "counter-revolutionaries" and nationalists

[13] For example, in a letter to his deputy Lazar Kaganovich from August 11, 1932, Stalin mentioned that the party district committees in about fifty districts in Ukraine had spoken out against state procurement quotas. He expressed his concerns that the Soviet government "could lose Ukraine" (Davies et al., eds, 2003). The evolving national procurement targets are reported by Davies and Wheatcroft (2009): p. 478, Table 20. Systematic data on the adjustments of the procurement targets are not available at the regional level.

[14] On May, 6 1933, during the peak famine mortality, Stalin wrote to Sholokhov, a famous writer originally from the Don region, that peasants "sabotaged" his policy and accused them of engaging in a "silent war" against the Soviet state (Murin, ed, 1997).

and fulfill procurement quotas (Danilov et al., eds, 1999-2006, Volume 3, Document 226).

Deaths from starvation peaked in the early months of 1933. In January 1933, Moscow ordered the closure of the borders between Ukraine and the Northern Caucus to prevent mass migration of peasants out of these areas (Danilov et al., eds, 1999-2006, vol.3, p.634-5). The government sent some food aid to rural areas, but the amount was insufficient. The best available data show that 176,000 tons of grain were sent to Ukraine in the spring and summer of 1933 (Davies and Wheatcroft, 2009, Table 23).[15] There were also other forms of famine relief, such as aid kitchens, medical assistance or housing. Most of these efforts were local, *ad hoc* and not systematically recorded. According to anecdotal accounts, rural famine victims often went to nearby urban areas to ask for food and some urban areas were known to have set up relief kitchens, while others were known to have aggressively expelled migrants back to rural areas. There are no systematic data about such relief efforts.

National mortality rates returned to trend in 1934, although some places took longer to recover. Total famine mortality estimates for the Soviet Union range from five to 10.8 million. Mortality was concentrated in rural areas though there were some accounts of famine in urban areas.[16]

There are no systematic data on ethnic-specific mortality rates in the Soviet Union for the period of our study. One way to approximate ethnic Ukrainian famine mortality is to use the most cited total famine death toll of seven million for the USSR (Conquest, 1986), and 2.6 million (Meslé et al., 2013) to 3.9 million (Rudnytskyi et al., 2015) for Ukraine. If famine deaths were equally distributed between ethnic Ukrainians (80% of Ukraine) and others ethnicities in Ukraine, and no ethnic Ukrainians died outside Ukraine, then ethnic Ukrainian deaths constitute 30% ($.8 \times 2.6/7 = .3$) to 45% ($.8 \times 3.9/7 = .45$) of the total famine deaths. Thus, ethnic Ukrainians, who were 21% of total Soviet population in 1926 constituted about 40% of all famine deaths.[17] Famine mortality rates in Ukraine were four to six times higher than rates in Russia.[18]

Many historians have argued that the strong resistance to agricultural collectivization from ethnic Ukrainians was an important reason for their systematic persecution.[19] Ukrainians had a strong group identity that included their own language and culture, which facilitated collective action. The Ukrainian communist

[15]This was around 2% of 1932 Ukrainian production according to the estimates that we will present in Table 1. There is no systematic data for most other regions.

[16]Conquest (1986) estimates total famine deaths to be 7 million. Davies and Wheatcroft (2009) estimates 5.5 to 6.5 million deaths. Ellman (2005) cites "'about eight and a half million' victims of famine and repression in 1930–33." Kondrashin (2008) gives a range between 5 and 7 million victims. Russian historical demographers estimate 7.2 to 10.8 million famine victims (Polyakov and Zhiromskaya, 2000). In 2008, the Russian State Duma postulated that within the territories of the Volga Region, the Central Black Earth Region, Caucasus, Ural, Crimea, Western Siberia, Kazakhstan, Ukraine and Belarus, the estimated famine death toll was 7 million people (State Duma, 2008). The difference in estimates are driven by data limitations, and potential underregistration of deaths during the famine. Estimates deriving excess famine deaths from a comparison of the pre-1926 and post-1937 Soviet population censuses suffer from the problem of the underregistration of infant mortality (See Davies and Wheatcroft, 2009, for a detailed discussion.).

[17]These estimates likely underestimate ethnic Ukrainian deaths because ethnic Ukrainian mortality was higher than those for other groups within Ukraine, and many Ukrainians who lived in other republics also died.

[18]If total famine deaths is seven million, and we subtract the deaths in Kazakhstan (1 to 1.5 million) and Ukraine (2.6 to 3.9 million), we are left with approximately 1.6 to 3.4 million deaths for Russia (assuming no famine mortality in other republics). This implies famine mortality rates of 14 to 30 per 1,000 for the 112 million residents of Russia. A similar calculation for Ukraine, yields a famine mortality rate of 81 to 122 per 1,000, which is around four to six times larger than the implied famine mortality rate in Russia.

[19]See, for example, Conquest (1986), Ellman (2007), Graziosi (2015), Mace (2004) and Snyder (2010) for variants of this argument.

party was the largest national branch of the Soviet Communist Party and viewed itself as representing the interests of ethnic Ukrainians across the Soviet Union, including those who lived outside the boundaries of the republic. Strong political opposition from Ukrainian nationalists during the Civil War was one of the main problems of the Bolshevik's "national question" (Graziosi, 2015). In other words, the Bolsheviks were concerned about controlling all peasants, but especially concerned about controlling the Ukrainian peasantry.

The Soviet government denied the existence of the 1932–33 famine in Ukraine and in the Soviet Union until the late 1980s.

3 Food Accounting

This section conducts a simple republic-level accounting exercise and documents that grain production in Ukraine in 1932 was sufficient for avoiding severe famine in Ukraine had there been no procurement. The results motivate the main analysis of ethnic Ukrainian bias in famine mortality.

Table 1 presents production, procurement, retention (production minus procurement) and compare retention to rural population food needs for each year for the years just before and after the famine, 1927-1939. These data cover the entire Soviet Union. Panel I examines Ukraine. Rows (1) and (2) report total and rural populations. These figures capture population for the beginning of each year. Thus, the population of 1933 does not account for the bulk of famine mortality which occurred during early 1933.[20] Row (3) reports production.

The main challenge for this exercise is that the aggregate production data are believed to have been exaggerated during the early years of collectivization because grain data were publicized and used as a marker for the success of Soviet economic policies and production had failed to meet expectations (Davies and Wheatcroft, 2009). Davies and Wheatcroft (2009) provide a range for the true aggregate production for the Soviet Union, but not for disaggregated geographic units. We follow the spirit of this earlier correction and approximate "true" production with data that were used only for internal purposes, classified until recently, of which the accuracy has never been disputed in academic studies, to construct corrected estimates for the region-level as well as the republic-level and the Soviet Union level production.

We use grain measures of procurement ratios (grain procurement stocks as a share of production) and procurement stocks to back out true production. Procurement stocks were directly observed and counted by the government and widely accepted as accurate. Production is an estimate. Procurement ratios were reported in Gosplan, a candid evaluation of the First Five-Year Plan shown only to the highest ranking Soviet officials. The report was classified until after the fall of the Soviet Union. The accuracy of the data for procurement ratios and procurement levels has not been disputed. Reassuringly, our corrected production measures are lower than officially reported production during the years when production is believed to have been exaggerated. At the Soviet Union level, our measures are comparable to the production estimates provided by Davies and Wheatcroft (2009). See the Appendix for more details.

[20]The population data are reported by different official sources. The missing years are due to the lack of population data for those years. See the Data Appendix. Later, for the regressions, we address these issues by examining mortality directly and by interpolating population for the missing years.

Row (4) presents grain procurement, which includes urban consumption, exports and national reserves. Production and procurement are reported in millions of tons.

Rural retention is the difference between production and procurement. Retention includes seed stock intended for cultivation since peasants consume seed stock during times of starvation. Row (5) reports retention in intuitive units of kilograms per capita per day units. To obtain these figures, we multiply the difference between production and procurement (rural retention in millions of tons) by 1,000 to convert the unit of measurement to kilograms, and divide by the rural population size and 365. Row (6) reports an alternative measure of rural retention that assumes that no grain is procured from rural Ukraine. Row (7) reports retention for all Ukrainians (urban and rural) if there is no procurement from Ukraine. This is simply production divided by the total population, converted in kilograms per capita per day units.

To understand whether retention is sufficient for avoiding famine, we calculate population food needs (row 8). We conservatively use official Soviet guidelines for maximum caloric needs for each gender and age group (e.g., heavy labor for prime-age adult males) as reported by Lositskij (1926) and Lositskij (1928). We use data on sex, age and urban-rural shares as reported by the 1926 census to adjust the guidelines by the demographic composition of each year.[21]

Row (3) shows that grain production in Ukraine declined from 23.2 to 16.8 million tons from 1930 to 1931, and further declined to 9.1 million tons in 1932. Row (4) shows that grain procurement was 7.7, 7.3 and 4.2 million tons in 1930, 1931 and 1932. The decline in procurement is consistent with the earlier historical discussion that the regime was unable to procure initial targets in the face of the production drop in 1932.

Several important facts emerge. First, grain procurement did not decline enough to prevent the famine. Row (5) shows that per capita grain retention in Ukraine declined from 1.7 kg/person/day after the harvest of 1930 to 1 kg/person/day after the harvest of 1931, and then only 0.5 kg/person/day after the harvest of 1932. 1932 grain retention was 30% below the estimated food needs of 0.778 kg/person/day presented (row 8).

Second, the intense famine in rural Ukraine could have been avoided, or at least, greatly moderated if there was no procurement from rural Ukraine. Row (6) shows that 1932 grain retention would have been 1.01 kg/person/day, enough to keep the rural population alive.

Third, lowering grain procurement to the level needed to avoid severe famine in rural Ukraine would not have necessitated famine in urban Ukraine. Row (7) shows that rural Ukraine produced enough food to sustain both urban and rural populations of the republic. Had there been no procurement out of Ukraine, 1932 grain production would have provided an average citizen of Ukraine with 0.79 kg/person/day, which is similar to the average food need of 0.78 kg/person/day.

Finally, not procuring food from Ukraine would not have necessitated famine elsewhere. Panel II examines all other republics of the USSR. Row (11) shows that urban and rural grain retention would have been 1 kg/person/day in 1932 if production in other republics were distributed equally across the population of the other republics. Retention would have been more than the 0.77 kg/person/day food need. See the Appendix

[21]The guidelines are generous and comparable to guidelines for other countries and international standards. For example, for a prime-age adult male, it is assumed that 3,750 calories are needed for heavy labor. We convert calories to kilograms of grain using calories per one kilogram of Russian grain estimated by Lositskij (1920).

for more descriptive statistics on production and procurement.

The aggregate grain production and procurement data show that enough food was produced in Ukraine to avoid a severe famine for its citizens and enough food was produced in the Soviet Union to avoid a famine in the Soviet Union. Thus, the drop in grain production in Ukraine, and all the contributors to the drop (e.g., weather, geography, earlier policies), cannot explain the famine. A necessary condition for famine in Ukraine is over procurement of grain from Ukraine.

The accounting results should be interpreted cautiously because the data are crude and measured with error. Specifically, there are several caveats to keep in mind for interpreting the accounting results. First, we focus on grain because it is the most important agricultural commodity for the Soviet economy. In 1928, 36% of all agricultural production was grain (Markevich and Harrison, 2011). It was also the main staple for consumption.[22] The availability of other food would strengthen our point that food production was sufficient for avoiding famine and that over procurement is a necessary condition for high famine mortality.

Second, our estimates do not take into account the possibility of the loss of foods in transportation or storage because of data limitations. Lositskij (1920) estimates such losses to be around 5%. However, our threshold for food need is very generous since the amount for heavy labor is much more than the amount needed for avoiding mortality. Thus, the point that famine would have been largely avoided if no food was taken out of Ukraine is unlikely to be overturned.

Third, note that the centrally planned procurement system comprised of two steps. The first step is to take food away from rural areas in each geographic region. This is what we report in the accounting exercise. The second step is to distribute a smaller amount of food back to the region. This is mostly food for urban areas, but also rural workers who were not involved in grain production (e.g., forestry workers). We do not include this in the accounting exercise (or other parts of the analysis) because the data are incomplete. Similarly, the accounting exercise focuses on centrally planned procurement and does not take into account *ad hoc* famine relief efforts (see Background Section). There are no systematic disaggregated data on such relief. Accounting for grain distributions back to rural areas and *ad hoc* relief will not overturn the point that enough food was produced to avoid famine in Ukraine.

Fourth, the estimates in this section do not take into account the facts that some of the population living in Ukraine are not ethnically Ukrainian, or the fact that a number of ethnic Ukrainians lived outside the republic of Ukraine in other Soviet regions.[23] We address this in the next section by inferring ethnic-specific mortality rates instead of examining republic-level mortality data.

Finally, interpreting our revised production estimates requires the assumption that changes in the procurement ratio reflects changes in production rather than changes in the ability to procure. We discuss the plausibility of this assumption in the Appendix. In the mortality analysis, we will address potential measurement error in production figures by using predicted instead of reported production.

[22] We do not have systematic disaggregated data for the production of other crops. Later, in the regression estimates, we will account for this by controlling for the suitability of cultivating other staple and cash crops.

[23] According to 1926 census, 80% of the total population and 87.5% of the rural population in Ukraine are ethnic Ukrainians.

4 Ukrainian Bias in Famine Mortality

This section uses province and district-level panel data to investigate whether there is Ukrainian bias in famine mortality.

4.1 Ukrainian Population Share and Famine Severity

Our main measure of famine severity is mortality (deaths per 1,000). At the region-level, the mortality data are available for each year in our sample for nineteen provinces from the three most populous Soviet republics: Belarus, Russia and Ukraine. The subsequent analyses will be restricted to these three republics, which include 84% of the 1926 Soviet population and 88% of the 1928 Soviet grain production. For brevity, we will refer to these three republics as the Soviet Union or the USSR in the remainder of the paper. As we discussed earlier, population data are not available for all years and we interpolate population and urban population for the missing years (see the Appendix). The average province in our sample has 6.5 million people in 1926. All data are mapped to the 1932 province borders.

Figure 1a plots mortality over time from 1923 to 1940 for Ukraine and for Belarus and Russia (the rest of our sample excluding Ukraine). It shows that during non-famine years, mortality rates are lower in Ukraine (18 per 1,000) than the other republics (22 per 1,000), and quite stable over time. However, mortality sharply increases during the famine. In Belarus and Russia, it increases in 1933 to approximately 30 per 1,000. In Ukraine, the increase begins slightly earlier in 1932 to approximately 22 per 1,000 and then spikes in 1933 to approximately 60 per 1,000. The earlier timing and larger magnitude of the total mortality increase in Ukraine are consistent with historical accounts that starvation began earlier in Ukraine and the greater intensity of the full-blown famine in 1933. To observe the granularity of the spatial variation in famine severity, see the province and district-level maps in the Appendix.[24]

The data on ethnic Ukrainian population share is reported by the 1926 Population Census, the last census before the famine. Ethnicity is self-reported. In 1926, ethnic Russians and Ukrainians constituted 53.1% and 21.3% of the Soviet Union, 57.2% and 23.1% of our sample, and 41.9% and 43.8% in provinces that central planners categorized as "grain surplus" regions. The next largest ethnic group was an order of a magnitude smaller: Belorussians were 3.2% of total Soviet Union population and 3.5% of our sample.[25]

Ukrainians mostly resided in Ukraine and in agriculturally productive regions, but many lived outside of Ukraine and the degree of agricultural productivity varied across regions where they lived. In the 1926 Census, 23.2 million Ukrainians lived in Ukraine and 7.9 million lived in the other two Soviet republics of our sample.[26] Since most non-Ukrainians in our sample are ethnic Russians, we will often refer to the

[24] Appendix Figure A.1c maps excess mortality at the district level. Appendix Figure A.1e maps the excess mortality after demeaning by province fixed effects. They show significant variation within provinces.

[25] The 1926 Population Census is commonly viewed as one of the highest quality Soviet censuses (Andreev et al., 1998). It is the last census before agricultural collectivization. Appendix Table A.1 Panel A lists the three largest ethnic groups in the entire Soviet Union, Panel B lists the three largest ethnic groups in our sample, and Panel C lists the three largest ethnic groups in the grain-surplus provinces of our sample.

[26] Appendix Figure A.1b maps the share of ethnic Ukrainians in the rural population for each province as reported in the 1926 Census. Grain-surplus regions are shaded in crosses. Appendix Figure A.1d maps Ukrainians at the district level. Appendix Figure A.1f maps Ukrainians at the district level after demeaning by province fixed effects. It shows that there is significant variation within provinces.

reference group in the regressions as "Russians" for brevity.

The following equation characterizes the baseline relationship between famine mortality and pre-famine ethnic Ukrainian population share.

$$mortality_{i,t+1} = \alpha + \beta Ukrainian_i \times Famine_t + \Gamma X_{it} + \eta_i + \delta_t + \varepsilon_{it}. \quad (1)$$

Mortality in province i during year $t+1$ is a function of: the interaction of the share of ethnic Ukrainians in the rural population of province i in 1926, $Ukrainian_i$, and a dummy variable that equals one in the famine year, $Famine_t$; a vector of additional controls, X_{it}; province fixed effects η_i; and year fixed effects δ_t. Since $Ukrainian_i$ is a time-invariant measure, the uninteracted term is absorbed by the province fixed effects. We define the famine dummy, $Famine_t$, to equal one in 1932 because 1933 was the year with the highest mortality rates and we assume that grain production in year t is used mostly to feed the population in year $t+1$. We adjust the standard errors for spatial correlation.[27]

The coefficient of interest is β, the difference in the correlation of Ukrainian share and mortality between famine and non-famine years. Conceptually, it captures excess (or abnormal) mortality that occurred during the famine. If ethnic Ukrainians died at higher rates during the famine, then $\beta > 0$.

To isolate Ukrainian bias in famine mortality rates, we control for other factors that may be associated with Ukrainian population share and mortality, X_{it}. The most important of these controls is per capita grain production. We control for both the uninteracted term as well as its interaction with the famine dummy variable to allow the relationship between mortality and production to vary between famine and non famine years. Recall the concern that official figures on grain production from the early 1930s are exaggerated. In the accounting exercise, we addressed this measurement error by revising official production estimates with previously classified data sources. However, the raw data we use for the correction are only available for a few years (1928–1933). Thus, to maximize the time span of the panel for the mortality analysis, the baseline regression instead controls for predicted grain, $PredictedGrain_{it}$.

$PredictedGrain_{it}$ is the measure of production predicted by exogenous factors such as weather and geography and can be interpreted as a parsimonious way of controlling for weather. We use pre-Soviet province-level grain production, monthly temperature and precipitation from Matsuura and Willmott (2014) and data for other inputs (e.g., log province area) to estimate the relationship between weather and grain for the years 1901 to 1915. We then use data for these variables from 1922 to 1940 to predict grain output.[28] The main concern for predicting production with weather is the possibility that the returns to inputs changed in the early Soviet era. We address this after we present the baseline estimates.

[27] We follow the recommendations by Colella et al. (2019) in adjusting for spatial correlation within 1,500 kilometers (the mean province width in our sample is 1,300 km). Note that one of the findings of this paper is that systematic bias against ethnic Ukrainians does not differ across administrative boundaries, and it is true across provinces as well as across districts within provinces. Thus, it makes sense to allow the correlation of the standard errors to gradually decay across space rather than cluster at the province level. Later, when we analyze much smaller geographic units – districts, we will estimate both standard errors that are spatially corrected and clustered at the district level.

Our results are qualitatively similar if the famine variable takes the value of one for 1931 and 1932. They are available upon request.

[28] See the Appendix Section B for details. The results are also similar if we directly control for weather instead of predicted grain (Appendix Table A.5).

To account for urban-rural differences in food policies and other factors that affect famine mortality, X_{it} also includes urban population share in province i during year t, and its interaction with $Famine_t$. We will show that the results are similar if we control for 1926 urban share (or lag urban share) interacted with the famine dummy.

Table 2 column (1) presents the baseline specification. The interaction coefficient of Ukrainian population share and the famine dummy variable is positive and statistically significant at the 1% level. The fact that we are controlling for predicted grain production means that higher famine mortality in ethnic Ukrainian areas was not due to different weather conditions. Taken literally, column (1) implies that for two provinces that experienced the same weather, and which have the same degree of urbanization, famine mortality rate was higher by 51 per 1,000 in a province with 100% Ukrainians than in a province with no Ukrainians.

The magnitude of the effect is very large since mean mortality rates are 21 per 1,000 during non-famine years and 31 per 1,000 during the famine. Another way to assess the magnitude is to examine the standardized coefficient, which is presented in italics. During the famine, increasing Ukrainian population share by one standard deviation would result in a 0.826 of standard deviation more in mortality relative to normal years.

The baseline control of predicted production conceptually holds production constant by holding weather and other important production inputs like total land area constant. However, if the relationship between these factors and production changes between the pre-Soviet era and the Soviet era, then the predicted estimates may be misleading. Specifically, one may be concerned that Soviet policy reduced the returns to these production inputs so that predicted production overstates true production in Ukrainian regions.

We address this in two ways. First, we alternatively control for the revised estimates of realized production that we used for the food accounting exercise in Table 1. The sample size is much smaller because the data we use to make the revisions are only available for a few years. However, the interaction coefficient is very similar to the baseline. Second, we return to our baseline and control for predicted grain as well as the policies that were most likely to have reduced agricultural productivity. In columns (3) and (4), we examine the sensitivity of the ethnic Ukrainian interaction coefficient to controlling for the two main policies that would have reduced the returns to normal agricultural inputs: *dekulakization* and the loss in livestock that occurred just prior to the famine. Approximately two million peasants were exiled to Siberia and other remote regions for actively resisting collectivization in the *dekulakization* campaign, amongst whom approximately 500,000 perished (Viola, 2007). *Kulaks* were often the relatively more productive peasants and their removal could have reduced the returns to the inputs we use to predict production.

Similarly, the loss of livestock could have had large negative effects on productivity. When peasants lost the property rights to their livestock, they responded by slaughtering, eating or simply neglecting the newly collectivized animals. Between 1929 and 1932, the number of horses declined by 42%, cattle by 40% (Viola, 1996, p. 70). Livestock was the main source of horsepower. The depletion of livestock also meant that the traditional means of avoiding famine – slaughtering and eating the animals – were unavailable to Soviet peasants.

We control for the number of kulak households exiled from each region in 1930-31 divided by the 1930 population (column 3) and the drop in per capita livestock between 1929 and July of 1931 (column 4). Since

these variables are time invariant, we control for their interactions with the famine indicator. The interactions of Ukrainian population share and the famine dummy variable are similar to the baseline. The estimates are similarly robust when controlling for alternative measures of the Soviet *dekulakization* campaign (Appendix Table A.5).

The results in columns (3)-(4) show that the main finding that areas with higher share of ethnic Ukrainians suffered higher famine mortality controlling for food production is unlikely to be confounded by systematic mis-measurement of production.

Column (5) controls for 1928 mortality rates interacted with the famine indicator. This addresses potential mechanical mean reversion in mortality rates. Again, the ethnic Ukrainian interaction coefficient is nearly identical to the baseline.

In specifying the baseline, we faced the usual tradeoff of under controlling (omitted variables bias) against over controlling. The baseline controls for province fixed effects because of concerns of omitted variables – e.g., mortality will vary across regions for reasons unrelated to Ukrainian bias or the famine. At the same time, one may be concerned that the inclusion province fixed effects over-controls and absorbs meaningful variation. We examine this concern by including the uninteracted Ukrainian effect instead of the province fixed effects. The ethnic Ukrainian interaction coefficient in column (6) is identical to the baseline in column (2), which means that, in practice, over-controlling is not a problem in the baseline. Interestingly, the uninteracted Ukrainian coefficient is -0.007 and statistically significant at the 1% level. This implies that in non-famine years, ethnic Ukrainian population share is *negatively* associated with mortality. It is only during the famine that mortality is *positively* associated with ethnic Ukrainian population share. The sum of the interaction and uninteracted coefficients presented at the bottom of the table is positive and statistically significant at the 1% level.

The lack of ethnic-specific mortality data has meant that past discussions of Ukrainian famine mortality have focused on deaths in Ukraine, which is one province in our sample. In 1926, 25% of ethnic Ukrainians in our sample lived in other provinces. Column (7) examines the Ukrainian-mortality gradient for these other Ukrainians. We re-estimate the baseline equation with a sample that excludes Ukraine. The main interaction coefficient is positive, large in magnitude and statistically significant at the 1% level. The standardized coefficient is similar to the baseline. Thus, famine mortality is systematically increasing in ethnic Ukrainian population share across all of the Soviet Union. Column (8) additionally excludes three other regions where agricultural productivity was particularly high and many Ukrainians lived (Lower Volga, North Caucasus, and West Siberia). The point estimates is similar to the baseline. The standardized coefficient is smaller because the variation in Ukrainian population share declines with these omissions.

Since the Ukrainian population was relatively concentrated, one may be concerned that our results are driven by outliers. The results in columns (7) and (8) goes against outliers or particularly influential observations driving our findings. Another way to address this concern is to conduct a random permutation test to demonstrate that our results are highly unlikely to be driven by coincidence. We permute Ukrainian population share across provinces (while preserving the mean and variance of the sample distribution). We then estimate the baseline equation using the randomly assigned Ukrainian population share. This is done for 10,000 iterations. Comparing the distribution of the coefficients from the permutations to the baseline

estimate (indicated by the vertical red line) in Appendix Figure A.3, we find that the probability that the latter is due to coincidence is 1% or less. We conduct a similar exercise where we permute 1933 mortality across provinces. Again, Appendix Figure A.3 shows that our findings are highly unlikely to be due to chance.

Recall that we interpolate the population data for years for which we do no have data to maximize coverage. Column (9) checks the sensitivity of our estimates to the interpolated data by using only population data from the 1926 census. The dependent variable is the number of deaths in province i and year t divided by total population in province i in 1926 (instead of year t), and we control for urban population share in 1926 (instead of yer t) and its interaction with famine. The point estimate and standardized coefficient are similar to the baseline. The robustness of the estimates in column (9) also alleviates any concerns about controlling for contemporaneous urban population, which may be endogenous to the famine, in the baseline.

The estimates in Table 2 provide strong evidence for the presence of systematic Ukrainian bias in famine mortality. To quantify the total contribution of Ukrainian bias to famine mortality, we conduct a simple back-of-the-envelope calculation using the baseline estimate of the association between rural ethnic Ukrainian population share and famine mortality: Table 2 column (1). The baseline regression model predicts that the number of deaths is on average 2.72 million in non-famine years, and 4.97 million in 1933.[29] The number of excess deaths due to the famine is the difference between mortality during famine and non-famine years: 2.26 million ($4.97 - 2.72 = 2.26$ million, with rounding error). We obtain the counterfactual famine mortality in a world with no Ukrainian bias by setting the interaction coefficient of ethnic Ukrainian population share and the famine dummy variable in equation (1) to zero. When we do this, predicted deaths in 1933 is 3.23 million. The number of famine deaths without bias against ethnic Ukrainians is the difference between this number and the number of deaths in non-famine years, 0.51 million ($3.23 - 2.72 = 0.51$ million). Thus, bias against Ukrainians accounts for 77% ($1 - 0.51/2.26 = 0.77$) of famine deaths in our sample. Since most non-Ukrainians in our sample are Russians, who suffered much lower famine mortality rates, our results imply that total famine mortality would have been 77% lower if ethnic Ukrainians died at similar rates as ethnic Russians.

We repeat the exercise for only Ukraine and find that bias against ethnic Ukrainians accounts for 92% ($1-0.12/1.51 = 0.92$) of famine deaths in Ukraine. The larger magnitude of the contribution for Ukrainian bias to famine mortality in Ukraine is due to Ukraine having a much higher share of ethnic Ukrainians.

For interpretation, note that Ukrainian bias in the mortality estimate includes bias in any policy that contributed to famine mortality. This includes centralized grain procurement, which we discussed in the previous section and will examine directly later in the paper. It also includes the distribution of other foods and famine relief that we do not observe in the data.

4.2 Dynamic Estimates

To investigate the dynamic patterns of the mortality-Ukrainian gradient, we estimate an equation similar to the baseline, except that we interact ethnic Ukrainian population share (and the other interacted controls) with dummy variables for each year instead of only 1932. Each interaction coefficient is the difference in

[29] This is comparable to the 4.81 million deaths in 1933 in the raw data.

mortality rate in year $t + 1$ between regions with 100% Ukrainian population share and regions with zero Ukrainian population share relative to the mortality difference in the reference year, 1923.

Figure 1b plots the interaction coefficients and their 95% confidence intervals and shows a sharp temporal pattern that goes against the concern that the baseline estimate is confounded by spurious correlations.[30] The correlation between Ukrainian population share and mortality becomes positive in 1932 and peaks in 1933. There is no correlation in other years.[31] This pattern is consistent with historical accounts of some starvation after the 1931 harvest, which then became a severe famine after the 1932 harvest. After 1933, regions with higher shares of Ukrainians had mortality rates similar to other regions.

4.3 Robustness

4.3.1 Sensitivity to Alternative Measurements

The first set of robustness exercises addresses the fact that our main dependent and explanatory variables may be measured with error and examines the sensitivity of our estimates to alternative measures.

Alternative Measures of Ukrainian Population Share and Mortality Rates The baseline measure of famine severity is total mortality rate because this is the standard measure in the famine literature, and in our context, this variable is available for a longer time horizon than rural or urban mortality rates.[32] We address the fact that most mortality was rural by controlling for urban population share and its interaction with the famine dummy variable. Here, we show that the results are robust to using alternative measures of mortality. Table 3 Panel A column (1) restates the baseline for comparison. Columns (2) and (3) replace total mortality with urban and rural mortality. The estimate for urban mortality in column (2) is small in magnitude and statistically insignificant. The estimate in column (3) for rural mortality is large and statistically significant at the 1% level.[33] The results confirm that higher famine mortality in regions with a larger share of ethnic Ukrainians was mostly a rural phenomenon.

We also show that our results are robust to using alternative measures of ethnic Ukrainian population share. Table 3 Panel A columns (4) to (5) show that our results are nearly identical if we use the total share of Ukrainians or urban Ukrainian population share. The interaction coefficient of the latter is larger than the baseline because the share of urban Ukrainians is smaller than the rural or total shares. The similarity of the standardized coefficients presented in italics in the table shows that the implied explanatory power of ethnic Ukrainian population on famine mortality is similar. Columns (6) and (7) use the share of people whose mother tongue is Ukrainian according to the 1926 and 1897 Population Censuses. The estimates are similar to the baseline.

Alternative Measures of Famine Severity We infer famine severity from the increase in mortality during the famine relative to other years. However, historians and demographers who have attempted to account

[30] The coefficients and their standard errors are presented in Appendix Table A.4

[31] Note that the confidence intervals are for the point estimates. The results in Table 2 column (6) show that the estimates for non-famine years are jointly statistically different from zero.

[32] Rural and urban mortality are available starting in 1926, while total mortality is available starting in 1923.

[33] Appendix Figure A.4a presents the dynamic estimates for rural mortality.

for total famine mortality note that many deaths were not registered during the famine (e.g., Davies and Wheatcroft, 2009). This could have been due to a breakdown of bureaucratic capacity or because the government or statistical reporting agency wanted to understate the severity of famine. If the undercounting of deaths is more severe in regions with more severe famine (i.e., regions with higher ethnic Ukrainian population share), then our estimates will understate Ukrainian bias.

We address this by examining two alternative measures of famine severity. The first is natality. These data also face the problem of under-registration, but the bias goes in the opposite direction as mortality. Live births should be decreasing in famine severity since starvation is negatively associated with the probability of pregnancy (and marriage), and is positively associated with the probability of miscarrying and stillbirths (Dyson and Ó Gráda, eds, 2002). If there is more under-registration of births in places with more severe famine (i.e., regions with higher ethnic Ukrainian population share), then our estimates will overstate Ukrainian bias. Thus, qualitatively similar findings between the natality and mortality estimates will lessen our concern that Ukrainian bias in famine mortality is an artifact of the under-registration of mortality.

Second, we use birth cohort size measured in the 1939 Census as the dependent variable. We use place of residence and age in 1939 to create a synthetic panel of province-specific birth cohort sizes. This measure does not face the problem of under-registration. However, it introduces another source of measurement error because of migration. Using birth cohort size to proxy for famine severity in a given province assumes that in 1939, people live in the same places as in 1932 and ignores the cross-province migration during the seven interim years. Nevertheless, it will be reassuring to find consistent results with this alternative data since they do not face the same measurement error problems as the official vital statistics data. The birth cohort size sample has more observations because provinces in the 1939 census are smaller than those we use in the earlier samples. We do not have annual population data for these smaller units and normalize cohort size with province population reported in the 1926 population census. We control for 1926 urban population share interacted with the famine instead of time varying urban population share and its interaction with the famine.[34]

Figures 1c and 1e plot natality rates and birth cohort sizes over time. The temporal patterns of the raw data are less sharp, but consistent with mortality. Average natality rates and birth cohort sizes begin declining around 1928 and reach the lowest levels in 1933 and 1934. They decline before mortality rates began to spike, and takes longer to recover back to trend than mortality rates. This is consistent with the fact that when food availability is low, food is generally allocated towards sustaining the existing population over giving birth, and that famine survivors need to recover their health before they can give birth. Note that the interpretation of natality and birth cohort sizes differ slightly. Lower natality changes reflect fewer births, while smaller birth cohorts reflect fewer births or lower survival rates.

Table 3 Panels B and C presents the natality and birth cohort size baseline estimates. The estimates are all negative and statistically significant at the 1% level. A back-of-the-envelope exercise using the natality and birth cohort size estimates show that Ukrainian bias explains 54% of missing births and 50% of missing survivors in 1939 in Ukraine, and 26% of missing births and 14% of missing survivors in 1939 in Belarus,

[34] Alternatively, we can normalize birth cohort size with the average size of all birth cohorts in each province as reported in the 1939 census. The results are nearly identical.

Russia and Ukraine.

Figures 1d and 1f plot the coefficients from the dynamic estimates.[35] The dynamic estimates for the two outcomes exhibit similar temporal patterns. For brevity, we focus the discussion on the estimates of birth cohort size. They show that provinces with high ethnic Ukrainian population share experienced the largest decline in births and child survival after the harvest of 1932. Prior to 1932, the interaction coefficients are negative but smaller in magnitude. The interaction coefficients begin to become statistically negative in 1926, which means that during the famine, those who were around eight years of age or younger were more likely to die from famine than older cohorts.

The estimates for natality and birth cohort sizes show that the main results are unlikely to be an artifact of mis-measurement in the mortality data.

4.3.2 Omitted Variables

The following exercises address concerns that pre-famine Ukrainian population share is correlated with omitted variables that influence famine severity.

Demographic Structure One may be concerned that young children are particularly vulnerable to famine, and higher mortality in ethnic Ukrainian areas may be due to differences in the demographic composition across regions. Table 3 column (8) controls for the population gender ratio and the share of individuals aged ten and younger (as reported by the 1926 Population Census), each interacted with the famine indicator. The ethnic Ukrainian interaction coefficient is 0.048 and statistically significant at the 1% level. The elderly are also more vulnerable during famines. In the Appendix, we show that our results are also robust to controlling for the share of the elderly, as well as a large number of alternative controls for the age and gender structure (see Appendix Table A.5).

Other Food Production We follow existing studies of the Soviet famine and focus on the production of grain, which was the main agricultural commodity and source of calories for the population. We do not have systematic disaggregated data for the production of other foods. If other food production is systematically different in Ukrainian areas, then the estimated mortality-Ukrainian gradient is biased. To address this, we control for the suitability of each province for the cultivation of other crops. Since suitability is mostly determined by geography, our first controls are crude: the interaction of latitude, longitude and the famine dummy (as well as the lower order interaction terms). Then, we control for the suitability for each important staple and crop – potato, wheat, rye, sugar beets, sunflowers, and flax – as predicted by the GAEZ model interacted with the famine dummy.[36] Table 3 columns (9)-(10) present the estimates. The Ukrainian interaction coefficient is robust.

Social Capital and Other Slow-Moving, Historical Factors A recent study by Buggle and Durante (2021) argues that social capital, which is higher in northern latitudes in Europe, takes a long time to form

[35] See Appendix Table A.4 for the coefficients and standard errors.

[36] Cotton is another important cash crop for the Soviet Union. It is mostly cultivated in regions outside of our sample.

and can play an important role for surviving famines. The fact that our results are robust to controlling for the interaction of latitude, longitude and the famine dummy suggests that they are not confounded by this slow-moving outcome of geography and history. We further investigate the role of Ukrainian-specific cultural norms or social networks by examining the 1892 famine. This was the last large famine in the Russian empire, during which approximately 500,000 people died. We obtain province-level mortality data from 1885 to 1913.[37] Table 3 Panel A column (11) presents the baseline specification for the 1892 famine and show that famine mortality is not associated with ethnic Ukrainian population share. Thus, the Ukrainian-mortality gradient is specific to the Soviet famine of 1932-33 and unlikely to be explained by slow-moving features of Ukrainian culture.[38]

We also directly control for historical factors that could affect famine severity. Table 4 Panel A controls for the interactions of various proxies of pre-Soviet regional wealth.[39] Panel B controls for proxies of cultural norms, institutions and inequality: the share of serfs in 1858 (three years before the abolition of serfdom), the shares of Catholics and Orthodox Christians (the two major religion groups in Ukraine) from the 1897 Population Census, the share of peasant households in repartition communes and the land Gini estimated from the 1905 Land Census, and peasant revolts per capita from 1895 to 1914. Our results are robust.[40]

Administrative Capacity This section addresses the concern that Ukrainian bias in famine mortality is an unintended consequence of differences in administrative capacity between areas with high Ukrainian population shares and areas with low Ukrainian population shares. The Bolsheviks may have strengthened their presence in Ukrainian areas because they received less support from Ukrainians, or anticipated more resistance. This means that policies will be implemented more zealously in Ukrainian areas and lead to higher famine mortality even if there was no Ukrainian bias in the intent of the policy. We address this by controlling for proxies of administrative capacity.

First, we control for the share of votes for the Bolshevik Party in the 1917 Constituency Assembly election, the first and only election until the end of the Bolshevik rule.[41] Table 5 column (1) restates the baseline for comparison purposes. Column (2) controls for the Bolshevik vote share in 1917 and the famine dummy variable. The interaction coefficient of interest is robust.

Next, we control for the size of the Communist Party in each province just before the famine: the

[37] We are grateful to Volha Charnysh for sharing 1885–1896 mortality and natality data from Charnysh (2022). The sample includes the fifty European provinces of the Russian Empire.

[38] Interestingly, Panel B column (11) shows that Ukrainian population share was positively associated with birth rates during the Tsarist famine. This is consistent with the fact that Ukrainians lived in more agriculturally productive regions and the conventional wisdom that absent state intervention (i.e., the Sen mechanism), famines are less severe in agricultural regions.

[39] These proxies are the nominal regional income per capita in 1897, real regional income per capita in 1897, regional labor productivity in 1897, regional rural labor productivity in 1897 (upper and lower bound estimates) from Markevich (2019), the value of agricultural equipment in 1910 and livestock in 1916 (from Castañeda Dower and Markevich, 2018).

[40] Living in a repartition commune required a more cooperative behavior (than in a hereditary commune), and according to the 1905 Land Census, repartition communes were less widespread in Ukrainian-populated regions than in Russian-populated regions. If the values of cooperation were transmitted down generations, this difference could contribute to the difference in mortality between the two ethnicities.

[41] Approximately 60% of the eligible voters turned out to vote. We follow Castañeda Dower and Markevich (2021) and use disaggregated district-level data on votes for the Bolsheviks from Protasov et al. (2014).

number of Communist Party Members (averaged over 1922, 1927 and 1931) per 1,000 individuals in each province. Party members were the key enforcers of state policy in the countryside and were responsible for grain procurement. Column (3) shows that the main result is unchanged when we control for this variable interacted with the famine dummy variable.

In column (4), we control for another measure of state capacity: the number of Party secretaries (at the province, district, city and, if the city was large, the borough level) who attended the 1930 Party Congress to vote formally for the policy of comprehensive collectivization. The Congress was a showcase of support for collectivization and the number of voting delegates can be interpreted as a proxy for the administrative capacity for implementing collectivization. Column (4) shows that our baseline is very robust to the interaction of this variable and the famine dummy variable.

The results in this section go against the concern that Ukrainian bias in famine mortality is due to spurious differences in administrative capacity.

4.4 District-Level Analysis

The district-level panel consists of two years: 1928 and 1933. Almost all of the data are manually collected from former Soviet archives. District-level mortality data are only available for Russia and Ukraine. Nevertheless, the increased granularity allows us to provide several additional pieces of evidence.

First, we examine whether famine mortality differs on the two sides of the Ukraine-Russia border. We define excess mortality as the difference between 1933 and 1928 mortality rates, and plot this excess mortality against the distance to the border between Ukraine and Russia, together with the fitted lines and their 95% confidence interval. Figure 4a plots excess mortality and shows that there is discrete decline in mortality rates as one crosses the border from Ukraine to Russia. This is consistent with anecdotal evidence that famine mortality was notably different across the border.[42] It is also consistent with the migration ban for Ukraine and the Northern Caucasus introduced by the Soviet government in January 1933 and conventional wisdom that movement of food and people between famine-stricken regions and better off regions was very limited. To see this, consider the hypothetical scenario of free mobility of food or people. In that case, we should not observe discrete changes in famine mortality rates along any administrative border.

Figure 4b plots the mortality residuals from a regression controlling for 1926 Ukrainian population share against distance to the border. It shows that the discrete change in famine mortality along the border disappears once we control for the ethnic Ukrainians rural population share of each district. This result emphasizes that Ukrainian bias was delineated along ethnic and not administrative lines. It is consistent with the earlier finding that our province-level results are qualitatively similar if we omit Ukraine (Table 2 column 7). The comparison of the two figures also shows how casual observations that do not account for other factors can be very misleading.

Second, the disaggregated data allow us to examine whether similar patterns exist across districts within provinces as across provinces. Soviet policies were centrally planned and implemented top-down by the bureaucracy. If the patterns we observe are driven by centrally planned procurement targets, then we would expect the association between famine mortality and Ukrainian population share across districts within

[42] See, for example, Applebaum (2017) Chapters 10 and 11 for survivors' recollections.

province to be similar to the association across provinces. Table 6 column (1) replicates the baseline specification with district and province-year fixed effects.[43] The province-year fixed effects isolate the within-province variation and control for factors that vary by province and year (e.g., regional political competition, leadership differences across provinces).[44] The results exhibit similar spatial patterns as the province-level estimates. This is consistent with the presence of a systematic and centrally planned policy.

Columns (2) to (10) show that the results are qualitatively similar when we subject the district-level estimates to the same sensitivity checks as the province-level estimates (to the extent that the data allow).[45] Similarly, the random permutation exercise with the district-level data shows that our results are unlikely to be due to coincidence. We alternately permute ethnic Ukrainian population share and 1933 mortality across districts (while preserving the mean and variance of the sample distribution). We then estimate the specification as in Table 6 column (1) using the randomly assigned ethnic Ukrainian population share or 1933 mortality. This is done for 10,000 iterations. Comparing the distribution of the coefficients from the permutations to the baseline estimate (indicated by the vertical red line) in Appendix Figure A.3, we find that the probability that the latter is due to coincidence is less than 1%.

5 Additional Results

This section examines whether there is Ukrainian bias in the main policy for food distribution – centrally planned grain procurement, investigates the long-run demographic effect of Ukrainian bias, and provides a speculative discussion on the motivation behind the Bolsheviks' bias against Ukrainians.

5.1 Realized Procurement and Retention

Table 7 documents the relationship between Ukrainian bias, centrally planned procurement, retention and famine mortality. The sample size is smaller than the main analysis because of data limitations in province-level procurement and production data. Column (1) examines the relationship between Ukrainian bias and procurement. The dependent variable is realized procurement as a share of realized production. We use our corrected grain production figures as the denominator. The estimate is similar to equation (1), except that we no longer control for predicted grain and its interaction with famine because grain production is the denominator of the dependent variable. The Ukrainian interaction coefficient is positive, 0.185, and statistically significant at the 1% level. Taken literally, it implies that all else equal, 1932 procurement from a province with 100% rural Ukrainian population share is 18.5 percentage-points higher than procurement from a province with no Ukrainians.

[43] Note that we use urbanization from 1926 and 1933 because urbanization at the district level is not available for 1928. We also use FAO GAEZ data base to construct a grain suitability index for each district because data limitations prevent us from predicting grain production at the district level.

[44] When using the district-level panel, we estimate standard errors that are adjusted for spatial correlation. We follow the recommendations by Colella et al. (2019) and adjust for spatial correlation within 400 kilometers (the mean district width in our sample is 76 km). The distance of 400 km delivers the largest (most conservative) standard errors. We can alternatively cluster the standard errors at the district level. We also cluster the standard errors at the district level. They are very similar and presented in brackets in the table.

[45] We do not observe age structure at the district level. Thus, we control for the gender ratio of the whole population to address potential differences in demographic structure.

Column (2) examines per capita retention as the dependent variable. It shows that all else equal, a province with 100% rural Ukrainian population share retained 1.085 kg/person/day less than a province with no Ukrainians. The Ukrainian interaction coefficient is positive and statistically significant at the 1% level. To examine the timing of Ukrainian bias in food retention, we replace the interaction of rural ethnic Ukrainian population share and the famine dummy with the interaction of rural ethnic Ukrainian population share and year dummy variables. 1928 is the omitted reference group. Figure 2 shows that Ukrainian population share is uncorrelated with food retention for most years, but negatively correlated during the famine.[46]

The estimates in columns (1) and (2) show that there was Ukrainian bias in centrally planned food procurement.

Columns (3) and (4) document the relationship between mortality and grain retention. In column (3), we regress mortality on realized grain retention and its squared term. Consistent with the positive and concave relationship between food consumption and mortality, we find that the coefficient for grain retention is negative and the coefficient for the squared term is positive. Both estimates are statistically significant at the 1% level. In column (4), we add the interactions of these variables with the famine dummy. The interactions are statistically zero, while the uninteracted terms are similar to column (3). This reflects the fact that the biological relationship between food consumption and mortality is stable over time.

The estimates support the interpretation that Ukrainian bias in Soviet policy contributed to high Ukrainian famine mortality: the Soviets procure more grain from Ukrainians, resulting in lower food retention for Ukrainians, which, in turn, results in higher mortality. To understand the importance of grain procurement policy as a channel for Ukrainian bias, we use this logic to conduct a simple quantification exercise.

First, we predict mortality given grain retention by estimating the relationship between mortality and grain retention.

$$mortality_{i,t+1} = F(retention_{it}) + \Gamma X_{it} + \eta_i + \delta_t + \varepsilon_{it}. \tag{2}$$

This is conceptually similar to the estimates shown in Table 7, but with a more flexible functional form and fewer controls (for parsimony and transparency). $F(retention_{it})$ is a flexible step function defined over 0.5 kilogram per person per day intervals of grain retention, X_{it} includes province characteristics that may affect mortality (e.g., urbanization), and η_i and δ_t are province and year fixed effects. We estimate this regression without a constant. Figure 3a plots the estimated step function. Lower retention is associated with higher mortality. [47] Applying the 1932 realized grain retention to the estimates from equation (2) implies 1.42 million deaths in Ukraine and 4.41 million deaths in the full sample.[48]

To predict grain retention for the counterfactual of no Ukrainian bias in the procurement policy, we use the estimates of the relationship between grain retention and the interaction of Ukrainian population share

[46] Appendix Table A.6 columns (7)-(9) reports the coefficients and standard errors.

[47] See Appendix Table A.6 for the coefficients and standard errors. The results are robust if in addition to urbanization, we control for urbanization interacted with the famine indicator and the official 1928 grain production interacted with the famine indicator.

[48] These numbers are slightly below the mortality numbers in the raw data: 4.81 and 1.86 million deaths for the full sample and Ukraine. This is because there are few observations with exceptionally high mortality in the data for which our estimated function under-predicts mortality.

and year dummy variable shown in Figure 2. We set the interaction of Ukrainian population share to zero to predict the counterfactual grain retention, and then apply the counterfactual grain retention to the estimates from equation (2): there are 3.84 million predicted deaths for the whole sample and 0.9 million for Ukraine.

Famine mortality with Ukrainian bias in grain retention is the difference between the number of deaths predicted by realized 1932 retention and the number in non-famine years: 0.9 million for Ukraine ($1.42 - 0.52 = 0.9$) and 1.7 million for the whole sample ($4.41 - 2.72 = 1.7$). Famine mortality with no bias in grain retention is the difference between the number of deaths predicted by the counterfactual and the number in non-famine years: 0.38 million for Ukraine ($0.9 - 0.52 = 0.38$) and 1.12 million for the whole sample ($3.84 - 2.72 = 1.12$).[49]

It follows that Ukrainian bias in grain retention explains 58% of excess deaths in Ukraine ($1 - .38/.9 = .58$) and 34% of all excess deaths in our sample ($1 - 1.12/1.7 = .34$). Since overall Ukrainian bias explains 77% and 92% of famine mortality in the whole sample and Ukraine, the estimates imply that approximately half of the total Ukrainian bias effect on famine mortality takes place through bias in food retention. We conduct a similar exercise for natality and find that Ukrainian bias in grain retention explains 13% of missing births in our sample and 26% of missing births in Ukraine.[50] As with mortality, the bias in grain retention accounts for half of the total contribution of Ukrainian bias to reduced natality during the famine.

The findings are consistent with the fact that grain procurement is the most important policy for food distribution, but also with the fact that there are other policies which contribute to food availability (e.g. procurement of other crops) and famine assistance (see Background).

5.2 Production and Procurement Targets

To understand whether Ukrainian bias in grain procurement was intentional, we examine whether procurement targets are associated with the share of ethnic Ukrainians in each regions. The First Five-Year Plan, published in 1928, laid out production and procurement targets for each province for the years of 1928 to 1933. Table 8 column (1) regresses per capita production targets on the rural share of ethnic Ukrainians. We control for year fixed effects to account for the fact that the planners assumed a high rate of growth in all regions, and officially reported 1928 per capita grain production, which was the measure used by planners to account for regional agricultural productivity. We do not control for province fixed effects because we have limited variation with only five years of data.

The coefficients for Ukrainians and grain 1928 are both positive and statistically significant. Thus, for two places with the same observed production in 1928, the one with more ethnic Ukrainians was expected to produce more. This could reflect the regime's desire to be harsher with Ukrainians or a belief that Ukrainian peasants were less productive than other peasants facing similar natural conditions, and thus, all else equal, had more room to expand production.

To distinguish between the two explanations, column (2) examines per capita grain procurement targets as the dependent variable, while controlling for production targets, which captures differences in perceived

[49] Recall that the baseline regression model in Table 2 column (1) predicts that the number of deaths is on average 2.72 million in non-famine years.

[50] Figure 3b shows a similar estimate for natality. As expected, the relationship is increasing.

production capacities across regions. We find that for two provinces assigned the same grain production target, the one with more Ukrainians is assigned a higher grain procurement target. Column (3) shows a similar pattern when we examine procurement as a share of production as the dependent variable. The positive Ukrainian coefficient means that the regime intended to take more grain from Ukrainian areas after conditioning for factors such as Ukrainian production capacity. In other words, because we control for grain production targets, the Ukrainian bias in grain procurement targets cannot be driven by a belief that Ukrainian grain production had more room to grow.

In column (4), we replace procurement ratios with per capita grain retention targets (the difference between the reported production and procurement targets) as the dependent variable. The coefficient for ethnic Ukrainian population share is -0.275 and statistically significant at the 1% level. This means that for two places facing the same production targets, central planners intended for the one with 100% ethnic Ukrainians to have 0.275 kg of grain less per capita per day than the one with no Ukrainians. The magnitude of the discrepancy is sizable: it is approximately one-third of the official Soviet food requirement for heavy labor (see Section 3).

The results in this section show that as early as 1928, the regime had planned for Ukrainian areas to retain less grain than for other regions with the same level of grain production.

There are two important caveats to interpreting the target data. First, the Five Year Plan published aggregate targets for every year during 1928-33, but province-level targets for only the two end years of 1928 and 1933. We back out the province-level targets for the interim years by assuming that central planners set the targets for each province following the same formula used for setting the targets for the entire Soviet Union. See the Appendix for details. Second, the target data do not prove that the regime intended for Ukrainians to die from famine because the grain production targets are much higher than actual grain production for all regions. The relationship between mortality and available food is concave and the degree of unequal grain retention implied by the regression results would not result in famine mortality at high levels of grain retention.

Related to this point is the fact that our results do not indicate the timing of the decision for allowing so many Ukrainian famine deaths. For example, in one extreme scenario, the regime may have aimed to reduce the Ukrainian population size *ex ante* and set procurement targets in 1928 that would lead to high Ukrainian famine mortality. This assumes that the regime knew that the production targets were unattainable (i.e., that collectivization would fail) but planned to enforce the procurement targets. Alternatively, the regime may have aimed to penalize Ukrainians for being more troublesome without the intention to cause famine. This would be consistent with policies that penalized disloyal ethnic groups that came to be practiced by Stalin in the late 1930s and 1940s (Polyan, 2001). In this scenario, the regime intended for Ukrainians to retain less food than other ethnic groups with the same level of production, but was sincerely over-optimistic about future production (i.e., the success of collectivization) and planned for the Ukrainians to retain enough for subsistence. When actual production was lower than planned production for Ukrainians in 1931 and 1932, the regime still procured and caused a severe famine instead of allowing Ukrainians to keep their production to minimize famine. The empirical results cannot be conclusive about the regime's objective in this sense. Both of the scenarios discussed here are consistent with Ukrainian bias in Soviet policy contributing to high

Ukrainian mortality.

5.3 Long-term Effects

The main analysis examines the causes of the famine and focuses on the short run effects of Ukrainian bias on famine mortality, natality and survival (of young children). This section examines the long-run demographic effects using population censuses on the territory of the former Soviet Union from 1897 to 2002. Each observation is a province in a census year. The data are harmonized to use 1932 province borders.

First, we examine the size of the population, which in the census, reflect mortality, fertility and other factors such as migration. We regress both log population and log rural population on the 1926 ethnic Ukrainian share interacted with (census) year dummy variables. All of the estimates in this section control for province and year fixed effects. Figure 7a plots the interaction coefficients from the two regressions.[51] They show that relative to areas with no Ukrainians, total and rural populations in Ukrainian areas decline in size immediately after the famine and recover by 1959. Next, we examine ethnic Ukrainians as a share of total population. Figure 7b shows that ethnic Ukrainian population share declined in areas with high pre-famine Ukrainian population shares. In these same places, ethnic Russian population share increased. The estimates show that Ukrainian bias during the famine had long-lasting impacts on ethnic composition in the stricken regions.

5.4 Motivation of Ukrainian Bias

The empirical analyses in this paper provide strong evidence that high Ukrainian famine mortality was largely due to bias against Ukrainians in Soviet policy. This section uses the empirical and historical evidence to provide a speculative discussion to help shed light on this important question and provide some clues for future research.

The empirical finding that Ukrainians were systematically repressed during the famine naturally raises the question of why. It seems counterintuitive to suffer the loss of so much productive labor. Yet, according to many historians, it is precisely because of the importance of Ukrainian regions to Soviet agriculture that Ukrainians were repressed. Recall from the Background discussion that the Bolsheviks needed to control agricultural production. They were generally concerned about peasant resistance, and had reasons to be particularly concerned about resistance from Ukrainians, who made up the largest ethnic group on agriculturally productive lands and had a history of resisting the Bolsheviks.

We provide two pieces of suggestive evidence that are consistent with this motivation. The first is to validate the claim that Ukrainians offered stronger resistance to Soviet agricultural policies than other ethnic groups. We collect data on de-classified secret police reports about peasant resistance to collectivization. The reports are from January 1931 to March 1932, just before the peak of the famine. They include three types of resistance. The first is the number of anti-Soviet "violent acts" per 1,000 people, which include

[51] Appendix Table A.7 columns (1)-(2) report the coefficients and their standard errors. Note that for 1937, we use revised estimates for total population and total rural population provided by Russian demographers (Bogoyavlensky, 2013) because the official figures are widely believed to have been inflated. The revised estimates are not available by ethnicity.

murders or attempted murders of local officials, arsons and the destruction of collective farm or state property. The second is the number of mass demonstrations in the countryside. The third is the number of episodes when anti-Soviet leaflets were distributed in the countryside.

For brevity, we examine the first principal component of the three indicators on a cross-section of provinces. The results are similar if we examine each measure separately. Because there are only nineteen provinces, we present the stylized fact in Figures 5a and 5b, which illustrate the correlation between collectivization and resistance in the nine provinces with Ukrainian population share below the sample median and in the ten provinces with Ukrainian population share above the median. Since both collectivization and resistance to it could have been associated with grain productivity or urbanization, we control for official 1928 grain production and urban population share. The correlation is steeper in provinces with higher ethnic Ukrainian population share, which supports the belief that, holding agricultural productivity constant, Ukrainian peasants were able to offer more resistance than other ethnic groups. The p-value for the statistical difference between the two slopes is 0.09.

The second piece of evidence is that the famine mortality-Ukrainian gradient is steeper in places that were more important for agricultural production. This implies that there was more intense targeting of Ukrainians in agriculturally productive places. We regress mortality on the triple interaction of Ukrainian population share, the famine dummy variable, and the importance of a region for rural economic production as perceived by the state (measured with official 1928 grain production). Table 9 shows that the triple interaction coefficient with 1928 grain production is positive and is statistically significant at the 1% level. In column (2), we control for a parsimonious measure of state capacity (the first principal component of the variables that we examined earlier in Table 5). The triple interaction coefficient with 1928 grain production is robust to the additional control, while the the triple interaction coefficient with the administrative capacity proxy is small in magnitude and statistically insignificant. Thus, the Ukrainian-mortality gradient is steeper in places that are more important for agriculture, but does not vary with these other factors that are unrelated to agriculture. Figure 6 plots the dynamic triple interaction coefficients. The timing is very sharp. The triple interaction is zero in all years except during the famine, when it spikes up. This goes against concerns of spurious correlations.[52]

Column (3) repeats the estimate with district-level data. We do not observe state capacity or grain production at the district level. Thus, we control for the interaction of suitability for grain production and the famine dummy instead of predicted grain production and its interaction with famine, and do not control for administrative capacity factors. The triple interaction of Ukrainian population share, 1928 grain production and the famine dummy is positive, large in magnitude and statistically significant at the 1% level. The results are consistent with the province-level findings.

These empirical facts support the political economic explanation that the regime repressed Ukrainians to control agriculture. There are several complementary variants of this explanation. The first is that Ukrainian bias is partly driven by the presence of information rigidities in the command economy. Meng et al. (2015) argue that Chinese famine mortality, which occurred in a centrally planned procurement system modeled after the Soviet one, was partly due to the government's inability to adjust procurement because it did

[52]The coefficients and their standard errors are presented in Appendix Table A.4.

not trust local information. Neither farmers nor bureaucrats are incentivized to report truthfully.[53] Thus, the government will not lower procurement until they can verify that the production drop is caused by exogenous factors (e.g., natural disasters) and not peasants shirking, or if the ensuing famine is too politically costly. Information rigidities can explain the Ukrainian bias in Soviet policy and famine mortality if the Bolsheviks were particularly distrustful of Ukrainians. The leadership could have been particularly hesitant to believe reports of low production. If central planners believed that Ukrainians were hiding more of their production prior to the famine, then distrust of Ukrainians can also explain why they faced higher procurement targets.[54] This explanation is consistent with the positive triple interaction effect of grain production in 1928, ethnic Ukrainian population share and the famine dummy on mortality if the regime believed that Ukrainians in agriculturally important regions were more likely to lie about production.

This theory of famine is consistent with the Bolsheviks' primary objective for repression being the need to control agricultural production; under-reporting production is a form of Ukrainian resistance that threatens the Bolsheviks. At the same time, it is unlikely that information rigidities in the Soviet centrally planned food distribution system are the only motivation for repressing Ukrainians for two reasons. First, the quantification of the procurement mechanism shows that Ukrainian bias in grain procurement explains around half of the total contribution of Ukrainian bias to famine. This leaves room for other drivers of bias. Second, there are important differences between the two famines. Most notably, there is no ethnic delineation in the Chinese famine, which occurred in the rural political power base of the Chinese Communist Party, including the home provinces of the top Party leaders. In contrast, by the time of the Ukrainian famine, there had been a history of political conflict between the Ukrainian peasantry and the Bolsheviks, whose political power base lie in urban areas. Thus, the drivers of famine are likely to differ between the two contexts.[55]

Another variation of the political economic hypothesis stems from the well-known insights of Horowitz (1985). The logic is as follows. Given the need to control agriculture, the regime may have targeted ethnic Ukrainians because it lacked more precise information on the likelihood of subversion. Thus, Ukrainian population share was used as a crude marker on which central planners conditioned policies. This explanation is consistent with the presence of a Ukrainian-mortality gradient during the famine.[56] It is consistent with the positive triple interaction effect of grain production in 1928, ethnic Ukrainian population share and the famine dummy on mortality if the regime believed that the Ukrainian-subversion gradient was steeper in

[53]Farmers are not the residual claimants of production and therefore incentivized to under report production or shirk. Local bureaucrats may over report production to curry favor with the Party or under-report to build a local power base. Both are bad for the regime. Over reporting can lead to over procurement and famine, which can lower future production and be politically destabilizing. Under reporting lowers revenues.

[54]Recall that Ukrainians face higher grain procurement targets, holding grain production targets constant. Thus, higher grain procurement targets cannot be explained by the belief that Ukrainians shirked more since expectations of shirking and other aspects of productivity are captured in the production grain targets.

[55]That different political motivations are at play in the two contexts is consistent with the difference in political response after the famine. The Soviet government denied the occurrence of the 1932-33 famine until the 1980s. The Chinese top leadership, in contrast, publicly acknowledged the occurrence of famine and the suffering of rural areas (albeit with understated official mortality rates) soon afterwards. See Meng et al. (2015) for a discussion.

[56]The Ukrainian-famine mortality gradient is still positive and statistically significant, although slightly smaller in magnitude than the baseline, if we control for resistance to collectivization interacted with the famine dummy in the baseline mortality regression. The result is not presented, but is available upon request. This implies that either there are other types of resistance correlated with Ukrainian population share not captured by our data, or that the Horowitz (1985) hypothesis is unlikely to explain all of Ukrainian bias.

agricultural regions.

The variants of this motivation discussed in this section are not mutually exclusive. Nor do they preclude other motivations for Ukrainian bias unrelated to controlling agriculture. The discussion in this section suggests that the Bolshevik's need to control agricultural production was likely to have played an important role. It is beyond the scope of this paper to be conclusive about the motivations for Ukrainian bias.

6 Conclusion

The deaths of approximately 2.1 to 3.15 million ethnic Ukrainians during the Great Soviet Famine is one of the most controversial demographic and economic disasters of the 20th century. The demographic consequences were huge and long lasting. In the 1926 and 1939 Population Censuses, ethnic Ukrainians in the Soviet Union declined from 31.2 million (21.3% of total population) to 28.1 million (16.5% of total population). In areas that the Bolshevik regime marked as important for grain production, Ukrainians were the largest ethnic group in 1926, at 43.8%, just ahead of Russians, at 41.9%. By 1939, the two groups switched places, with Russians being 48.1% and Ukrainians being 37.1% of the population in productive regions.

The empirical evidence in this study shows that the main culprit of Ukrainian famine mortality was Ukrainian bias in Soviet policy. Ukraine produced enough food to avoid severe famine. The famine occurred because the Soviet government took too much food away from Ukraine. The bias did not differ across the administrative boundaries of Russia and Ukraine, but was instead systematically implemented based on the ethnic share of Ukrainians in each province and district across the Soviet Union.

Our findings raise several questions for future research. First, it is important to better understand what policies beyond grain procurement led to high Ukrainian famine mortality. Second, we need to better understand the political-economic drivers of the regime's motivation for its bias against Ukrainians, and relatedly, the political-economic causes of the Great Soviet famine in general. The results and discussion in this paper provide some clues, but there are other possible explanations. Finally, we need to better understand the consequences of the famine. Researchers have documented that historical exposure to famine affect anti-Russian sentiments (Rozenas and Zhukov, 2019) and economic behavior (Korovkin and Makarin, forthcoming) in Ukraine today. Much more research is needed.

This paper focuses on the academic debate about the causes of Ukrainian famine. The causes has been equally, if not more, intensely discussed amongst policymakers and the public. Popular historians such as Applebaum (2017) and policymakers such as those in the Ukrainian Parliament (2006) argue that the famine was a genocide of the ethnic Ukrainian population. The European Parliament condemned it as a "crime against humanity" in 2008. In contrast, the Soviet government denied the occurrence of the famine until the 1980s. And although the Great Soviet Famine is now generally recognized in Russia and the lower house of the Russian Parliament condemned the USSR's "disregard for the lives of people" in 2008, the Ukrainian famine experience is still highly controversial and politicized. For example, as recently as October, 2022, the Russian state news agency reported that Russian officials in Russian-annexed Mariupol removed a commemorative memorial of the victims of the *Holodomor* in an effort to remove a symbol of

"political disinformation".[57]

These controversies emphasize the relevance of understanding the causes of this historical tragedy for today. It is our hope that the rigorous empirical evidence provided in this paper can establish a ground truth of the basic facts.

References

Acemoglu, Daron, Tarek Hassan, and James Robinson, "Social Structure and Development: A Legacy of the Holocaust in Russia," *The Quarterly Journal of Economics*, 2019, *126*, 895–946.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya, "Radio and the Rise of The Nazis in Prewar Germany," *The Quarterly Journal of Economics*, 07 2015, *130* (4), 1885–1939.

Alesina, Alberto and Eliana La Ferrara, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, September 2005, *43* (3), 762–800.

Allen, Robert, *Farm to Factory: A Reinterpretation of the Soviet Industrial Revolution*, Princeton: Princeton University Press, 2003.

Andreev, E.M., L.E. Darskij, and T.L. Kharkova, *Demographicheskaya Istroriya Rossii: 1927-1959*, Moscow: Informatika, 1998.

Applebaum, Anne, *Red Famine: Stalin's War on Ukraine*, New York: Penguin Press, 2017.

Bauer, Thomas K., Sebastian Braun, and Michael Kvasnicka, "The Economic Integration of Forced Migrants: Evidence for Post-War Germany," *The Economic Journal*, 2013, *123* (571), 998–1024.

Becker, Sascha O., Irena Grosfeld, Pauline Grosjean, Nico Voigtlander, and Ekaterina Zhuravskaya, "Forced Migration and Human Capital: Evidence from Post-WWII Population Transfers," *American Economic Review*, May 2020, *110* (5), 1430–63.

Berelowitch, A. and V. Danilov, eds, *Les campagnes Soviétiques vues par la Tchéka–O.G.P.U.–N.K.V.D. 1918–1939. Recueil de documents et matériaux en 4 volumes. Sovetskaya derevnya glazami VChK–OGPU–NKVD. 1918–1939 gg. Dokumenty i materialy v 4 tomakh*, Moscow: Rosspen, 2000-2012.

Bogoyavlensky, Dmitry, "O pripiskakh v perepisi 1939 g [On the centralized additions in the 1939 census]," *Demoscope.ru*, 2013.

Buggle, Johannes and Stephen Nafziger, "The Long-Run Consequences of Labor Coercion: Evidence from Russian Serfdom," *The Review of Economics and Statistics*, 2019, *87*, 217–234.

Buggle, Johannes C and Ruben Durante, "Climate Risk, Cooperation and the Co-Evolution of Culture and Institutions," *The Economic Journal*, 01 2021, *131* (637), 1947–1987.

Burgess, Robin and Dave Donaldson, "Railroads and the Demise of Famine in Colonial India," Working Paper, MIT 2017.

[57] See RIA Novosti channel in telegram, October 19, 2022, https://t.me/rian_ru/182454. Accessed October 24, 2022.

Castañeda Dower, Paul and Andrei Markevich, "Labor Misallocation and Mass Mobilization: Russian Agriculture during the Great War," *The Review of Economics and Statistics*, 2018, *100(2)*, 245–59.

_ and _ , "Did Industrialization Increase Support for the Radical Left? Evidence from the 1917 Russian Revolution," *Working paper*, 2021.

_ , Evgeny Finkel, Scott Gehlbach, and Steven Nafziger, "Collective Action and Representation in Autocracies: Evidence from Russia's Great Reforms," *American Political Science Review*, 2018, *112* (1), 125–47.

Charnysh, Volha, "Explaining Out-Group Bias in Weak States: Religion and Legibility in the 1891/1892 Russian Famine," *World Politics*, 2022, *74* (2), 205–248.

Cheremukhin, Anton, Mikhail Golosov, Sergei Guriev, and Aleh Tsyvinski, "The Industrialization and Economic Development of Russia through the Lens of a Neoclassical Growth Model," *The Review of Economic Studies*, 2017, *84* (2), 613–649.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig, "Inference with Arbitrary Clustering," 2019.

Conquest, Robert, *The Harvest of Sorrow: Soviet Collectivization and the Terror-Famine*, Oxford, UK: Oxford University Press, 1986.

Danilov, Victor, Robert Manning, and Lynne Viola, eds, *Tragediya Sovetskoj Derevni. Kollektivizatsiya i raskulachivanie. Dokumenti i materialy v 5 tomakh, 1927-1939*, Rosspen, 1999-2006.

Davies, R. W., Oleg V. Khlevniuk, E. A. Rees, Liudmila P. Kosheleva, and Larisa A. Rogovaya, eds, *The Stalin-Kaganovich Correspondence, 1931-36*, New Haven: Yale University Press, 2003.

Davies, Robert W. and Stephen G. Wheatcroft, *The Years of Hunger: Soviet Agriculture, 1931-1933. Second edition.*, Vol. 5 of *The Industrialization of Soviet Russia*, Basingstoke, UK: Palgrave Macmillan, 2009.

DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya, "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia," *American Economic Journal: Applied Economics*, 2014, *6* (3), 103–132.

Dyson, Tim and Cormac Ó Gráda, eds, *Famine Demography: Perspectives from the Past and Present*, Oxford, UK: Oxford University Press, 2002.

Egorov, Georgy, Ruben Enikolopov, Alexey Makarin, and Maria Petrova, "Divided We Stay Home: Social Distancing and Ethnic Diversity," NBER Working Papers 27277, National Bureau of Economic Research, Inc May 2020.

Ellman, Michael, "The Role of Leadership Perceptions and of Intent in the Soviet Famine of 1931-1934," *Europe-Asia Studies*, 09 2005, *57* (6), 823–841.

_ , "Stalin and the Soviet famine of 1932-33 Revisited," *Europe-Asia Studies*, 2007, *59* (4), 663–693.

Esteban, Joan, Laura Mayoral, and Debraj Ray, "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, June 2012, *102* (4), 1310–42.

Finkel, E. and S. Gehlbach, *Reform and Rebellion in Weak States* Elements in Political Economy, Cambridge University Press, 2020.

Gokmen, Gunes and Dmitrii Kofanov, "Career incentives in political hierarchy: evidence from Imperial Russia," *European Review of Economic History*, 2020, *24(2)*, 264–287.

Gorodnichenko, Yuriy and Gerard Roland, "Culture, Institutions, and the Wealth of Nations," *The Review of Economics and Statistics*, July 2017, *99* (3), 402–416.

Graziosi, Andrea, "The uses of hunger Stalin's solution of the peasant and national questions in Soviet Ukraine, 1932 to 1933," in Lubomyr Luciuk Declan Curran and Andrew G. Newby, eds., *Famines in European History. The Last Great European Famines Reconsidered*, Routledge, 2015, pp. 223–260.

Grosfeld, Irena, Alexander Rodnyansky, and Ekaterina Zhuravskaya, "Persistent Antimarket Culture: A Legacy of the Pale of Settlement after the Holocaust," *American Economic Journal: Economic Policy*, August 2013, *5* (3), 189–226.

_ , Seyhun Orcan Sakalli, and Ekaterina Zhuravskaya, "Middleman Minorities and Ethnic Violence: Anti-Jewish Pogroms in the Russian Empire," *The Review of Economic Studies*, 01 2020, *87* (1), 289–342.

Horowitz, Donald L., *Ethnic groups in conflict*, Berkeley: University of California Press, 1985.

Kondrashin, Viktor, *Golod 1932-1933 godov: Tragediya Sovetskoj derevni*, Moscow: Rosspen, 2008.

Korovkin, Vasily and Alexey Makarin, "Conflict and Inter-Group Trade: Evidence from the 2014 Russia-Ukraine Crisis," *The American Economic Review*, forthcoming.

Kotkin, Stephen, *Stalin. Waiting for Hitler, 1929-1941*, New York: Penguin Press, 2017.

Kuromiya, Hiroaki, "The Soviet Famine of 1932-1933 Reconsidered," *Europe-Asia Studies*, 2008, *60* (4), 663–675.

Li, Wei and Dennis Tao Yang, "The Great Leap Forward: Anatomy of a Central Planning Disaster," *Journal of Political Economy*, 2005, *113* (4), 840–877.

Lositskij, A., "Pishchevoe znachenie sovremennikh produktov pitaniya [The nutritional value of modern foods]," *Vestnik statistiki*, 1920, *1920* (1-4), 102–121.

_ , *Sostoyanie pitaniya gorodskogo naseleniya SSSR, 1919-1924 [The State of Nutrition of the Urban Population of the USSR, 1919-1924]*, TsSU, 1926.

_ , *Sostoyanie pitaniya selskogo naseleniya SSSR v 1920-1924 [The State of Nutrition of the Rural Population of the USSR in 1920-1924]*, TsSU, 1928.

Mace, James E., "Soviet Man-Made Famine in Ukraine," in William S. Parsons Totten Samuel and Israel W. Charny, eds., *Century of Genocide: Critical Essays and Eyewitness Accounts.*, Routledge, 2004.

Markevich, Andrei, "A Regional Perspective on the Economic Development of the late Russian Empire," *Working paper*, 2019.

_ **and Ekaterina Zhuravskaya**, "The Economic Effects of the Abolition of Serfdom: Evidence from the Russian Empire," *American Economic Review*, April 2018, *108* (4-5), 1074–1117.

_ **and Mark Harrison**, "Great War, Civil War, and Recovery: Russia's National Income, 1913 to 1928," *Journal of Economic History*, 2011, *71(3)*, 672–703.

Martin, Terry, *The Affirmative Action Empire: Nations and Nationalism in the Soviet Union, 1923-1939*, Cornell University Press, 2001.

Matsuura, Kenji and Cort Willmott, "Terrestrial Air Temperature and Precipitation: 1900-2014 Gridded Monthly Time Series (Version 4.01)," *National Oceanic and Atmospheric Administration*, 2014.

Meng, Xin, Nancy Qian, and Pierre Yared, "The Institutional Causes of China's Great Famine, 1959-1961," *The Review of Economic Studies*, 2015, *82* (4), 1568–1611.

Meslé, France, Jacques Vallin, and Evgeny Andreev, "Demographic Consequences of the Great Famine: Then and Now," in Andrea Graziosi, Lubomyr A. Hajda, and Halyna Hryn, eds., *After the Holodomor: The Enduring Impact of the Great Famine on Ukraine*, Harvard University Press, 2013.

Montalvo, Josz G. and Marta Reynal-Querol, "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, June 2005, *95* (3), 796–816.

Murin, Yuri, ed., *Pisatel' i vozhd': perepiska M.A. Sholokhova s I.V. Stalinym 1931-1950 gody : sbornik dokumentov iz lichnogo arkhiva I.V. Stalina*, Raritet, 1997.

Naumenko, Natalya, "The Political Economy of Famine: the Ukrainian Famine of 1933," *The Journal of Economic History*, 2021, *81* (1), 156–197.

Ó Gráda, Cormac, *Famine: A Short History*, Princeton, NJ: Princeton University Press, 2009.

Petrova, Maria and David Yanagizawa-Drott, "Media Persuasion, Ethnic Hatred, and Mass Violence: a Brief Overview of Recent Research Advances," in Charles Anderton and Jurgen Brauer, eds., *Economics of Genocide, Mass Killing, and Their Prevention*, Oxford: Oxford University Press, 2016, chapter 12.

Polyakov, Yu.A. and V.B. Zhiromskaya, *Naselenie Rossii v XX veke: istoricheskie ocherky. Vol. 1 1900-1939*, Rosspen, 2000.

Polyan, Pavel, *Ne po svoej vole... Istoriya i geografiya prinuditelnikh migratsij v SSSR*, Moscow: O.G.I, 2001.

Protasov, L.G., A.V. Repnikov V.V. Zhuravlev, and V.V. Shelokhaev, *Vserossijskoe uchreditelnoe sobranie. Entsiklopedia. (All-Russian Constitutional Assembly. Encyclopedia)*, Moscow: Rosspen, 2014.

Roland, Gerard, "The Long-Run Weight of Communism or the Weight of Long-Run History?," WIDER Working Paper Series wp-2010-083, World Institute for Development Economic Research (UNU-WIDER) 2010.

Rozenas, Arturas and Yuri M. Zhukov, "Mass Repression and Political Loyalty: Evidence from Stalin's 'Terror by Hunger'," *American Political Science Review*, 2019, *113* (2), 569–583.

Rudnytskyi, Omelian, Natalia Levchuk, Oleh Wolowyna, Pavlo Shevchuk, and Alla Kovbasiuk, "Demography of a man-made human catastrophe: The case of massive famine in Ukraine 1932-1933," *Canadian Studies in Population*, 2015, *42* (1-2), 53–80.

Sen, Amartya, *Poverty and Famines. An Essay on Entitlement and Deprivation*, Oxford, UK: Oxford University Press, 1981.

Snyder, Timothy, *Bloodlands: Europe Between Hitler and Stalin*, New York City, NY: Basic Books, 2010.

Stalin, Iosif, "Concerning the National Question in Yugoslavia," *Bol'shevik*, 15 April 1925, *7*.

State Duma, "Russian: Postanovleniye GD FS RF ot 02.04.2008 N 262-5 GD O Zayavlenii Gosudarstvennoy Dumy Federal'nogo Sobraniya Rossiyskoy Federatsii "Pamyati zhertv goloda 30-kh godov na territorii SSSR"," 2008.

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

Tauger, Mark B., "The 1932 Harvest and the Famine of 1933," *Slavic Review*, 1991, *50* (1), 70–89.

Ukrainian Parliament, "About 1932-33 Holodomor in Ukraini," *Vedomosti Verkhovnoj radi Ukraini*, 2006, *50* (504).

Viola, Lynne, *Peasant Rebels Under Stalin: Collectivization and the Culture of Peasant Resistance*, Oxford, UK: Oxford University Press, 1996.

_ , *The Unknown Gulag: The Lost World of Stalin's Special Settlements*, Oxford: Oxford University Press, 2007.

Zhuravskaya, Ekaterina, Sergei Guriev, and Andrei Markevich, "New Russian Economic History," *The Journal of Economic Literature*, forthcoming.

Table 1: Food Availability

	1927	1928	1929	1930	1931	1932	1933	1937	1939
	I. Ukraine								
(1) Total population (mln)	29.0	29.6	30.3	30.8	31.3	31.7	31.9	28.4	29.6
(2) Rural population (mln)	23.6	24.6	24.9	25.1	25.0	24.8	25.0	18.8	18.7
(3) Production (mln tons)	.	14.9	18.7	23.2	16.8	9.1	17.7	.	.
(4) Procurement (mln tons)	4.0	2.0	5.3	7.7	7.3	4.2	6.1	.	.
(5) Rural retention (kg/person/day)	.	1.44	1.48	1.69	1.04	0.54	1.28	.	.
(6) Rural Retention, no procurement (kg/person/day)	.	1.66	2.06	2.53	1.84	1.01	1.95	.	.
(7) Rural and urban retention, no procurement (kg/person/day)	.	1.38	1.70	2.06	1.47	0.79	1.52	.	.
(8) Food needs for heavy labor (kg/person/day)	0.78	0.79	0.78	0.78	0.78	0.78	0.78	0.76	0.75
	II. USSR – no Ukraine								
(9) Total population (mln)	118.0	120.8	124.0	126.7	129.3	131.5	133.8	133.7	135.9
(10) Production (mln tons)	.	58.5	53.1	60.1	47.8	48.0	57.3	.	.
(11) Rural and urban retention, no procurement (kg/person/day)	.	1.33	1.17	1.30	1.01	1.00	1.17	.	.
(12) Food needs for heavy labor (kg/person/day)	0.78	0.78	0.78	0.78	0.78	0.77	0.77	0.76	0.76

Notes: Data for population and procurement are official statistics. Production is revised by the authors using archival sources. See the text and Appendix for details. Retention is the difference between production and procurement. Food needs are calculated by the authors and take into account the demographic composition (e.g., age, gender, rural/urban) as reported in the population censuses. They are based on official guidelines for maximum caloric needs for each group as reported by Lositskij (1926, 1928). Panel I includes Ukraine. Panel II includes all other republics in the USSR.

Table 2: Famine Mortality and Ethnic Ukrainian Population Share – Province-level Estimates

	Dependent Variable: Mortality in Year t+1								
	Baseline	Control for Revised Realized Grain Production				Omit Province FE	Omit Ukraine	Omit Ukraine, Lower Volga, North Caucasus, West Siberia	Use 1926 Census Population Measures
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. var. mean.	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.023*
Dep. var. mean. in 1933	*0.031*	*0.031*	*0.031*	*0.031*	*0.031*	*0.031*	*0.029*	*0.026*	*0.033*
Ukrainians × Famine [1]	0.051***	0.050***	0.054***	0.053***	0.050***	0.051***	0.086***	0.059***	0.056***
	(0.006)	(0.002)	(0.005)	(0.005)	(0.006)	(0.005)	(0.007)	(0.006)	(0.005)
Standardized Coef.	*0.826*	*0.814*	*0.870*	*0.858*	*0.819*	*0.831*	*0.840*	*0.410*	*0.968*
Ukrainians [2]						-0.007***			
						(0.002)			
Grain	0.0002	0.001	0.0002	0.0002	0.0002	-0.0002	0.0003**	0.0007***	0.0001
	(0.0001)	(0.001)	(0.0002)	(0.0002)	(0.0001)	(0.0004)	(0.0001)	(0.0002)	(0.0002)
	0.0004	0.007**	-0.003	-0.001	0.0004	-0.0001	-0.003	-0.002	-0.001
Grain × Famine	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.002)
Standardized Coef.	*0.023*	0.275	*-0.137*	*-0.067*	*0.021*	*-0.005*	*-0.183*	*-0.113*	*-0.056*
Kulaks × Famine			1.805						
			(1.367)						
Livestock Change × Famine				0.016					
				(0.015)					
Mortality 1928 × Famine					-0.087				
					(0.485)				
Urbanization	-0.005	0.016	-0.006	-0.006	-0.005	-0.011**	-0.004	-0.004	
	(0.007)	(0.020)	(0.007)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)	
Urbanization × Famine	0.003	0.011	-0.004	0.002	0.003	0.006	-0.003	-0.005	-0.005
	(0.007)	(0.009)	(0.006)	(0.006)	(0.007)	(0.008)	(0.007)	(0.005)	(0.005)
Observations	337	107	337	337	337	337	319	268	337
R-squared	0.776	0.838	0.791	0.784	0.777	0.420	0.758	0.808	0.765
[1] + [2]: Coef.						0.044			
p-val.						<0.001			

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Mortality is the number of deaths each year and province divided by the total population each year and province in columns (1)-(8); and the number of deaths in each year and province divided by the province's 1926 population in column (9). Ukrainians is the share of ethnic Ukrainians in the rural population. Famine is an indicator that equals one in 1932 and zero otherwise. Columns (1) and (3)-(9) control for per capita grain production predicted by exogenous factors. Column (2) controls for grain production estimates revised by the authors using archival data (see text). Urbanization is the urban population share. Column (3) controls for the number of kulak households exiled during 1930-31 per 1930 population interacted with the famine indicator. Column (4) controls for the drop in livestock (horses and cattle) per capita between 1929 and 1931 interacted with the famine indicator. Column (5) controls for the 1928 mortality rate interacted with the famine indicator. Column (9) controls for predicted grain per 1926 population and 1926 urban population share. All regressions control for province and year FE. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Famine Severity and Ethnic Ukrainian Population Share— Robustness to Alternative Measures and Additional Controls

	Baseline (1)	Urban Mortality/ Natality/ Cohort (2)	Rural Mortality/ Natality/ Cohort (3)	Total Ukrainians (4)	Urban Ukrainians (5)	Mother Tongue Ukrainian 1926 (6)	Mother Tongue Ukrainian 1897 (7)	Control for Demographic Structure × Famine (8)	Control for Latitude × Longitude × Famine (9)	Control for Famine × Suitability for: Potato, Sugarbeet, Sunflower, Wheat, Rye, Flax (10)	1892 Famine (11)
A. Dependent Variable: Mortality in Year t+1											
Dep. var. mean.	*0.022*	*0.021*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.032*
Ukrainians × Famine	0.051***	0.013	0.063***	0.055***	0.090***	0.056***	0.058***	0.048***	0.059***	0.056***	-0.0002
	(0.006)	(0.008)	(0.005)	(0.006)	(0.010)	(0.008)	(0.007)	(0.007)	(0.004)	(0.004)	(0.003)
Standardized Coef.	*0.826*	*0.252*	*0.869*	*0.820*	*0.785*	*0.747*	*0.820*	*0.781*	*0.965*	*0.913*	*-0.004*
Observations	337	285	285	337	337	337	337	337	337	337	1,297
R-squared	0.776	0.746	0.787	0.776	0.771	0.765	0.781	0.783	0.806	0.806	0.864
B. Dependent Variable: Natality in Year t+1											
Dep. var. mean.	*0.039*	*0.008*	*0.030*	*0.039*	*0.039*	*0.039*	*0.039*	*0.039*	*0.039*	*0.039*	*0.047*
Ukrainians × Famine	-0.014***	-0.002***	-0.012***	-0.015***	-0.025***	-0.015***	-0.016***	-0.015***	-0.014***	-0.009***	0.004**
	(0.003)	(0.0004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.005)	(0.002)	(0.003)	(0.002)
Standardized Coef.	*-0.432*	*-0.134*	*-0.350*	*-0.429*	*-0.409*	*-0.381*	*-0.424*	*-0.469*	*-0.419*	*-0.285*	*0.102*
Observations	337	285	285	337	337	337	337	337	337	337	1,297
R-squared	0.820	0.940	0.902	0.820	0.820	0.819	0.820	0.820	0.823	0.822	0.934
C. Dependent Variable: Birth Cohort Size in Year t+1/Total 1939 Population											
Dep. var. mean.	*0.024*	*0.024*	*0.024*	*0.024*	*0.024*	*0.024*	*0.024*	*0.024*	*0.024*	*0.024*	
Ukrainians × Famine	-0.009***	-0.009***	-0.009***	-0.010***	-0.016***	-0.010***	-0.010***	-0.010***	-0.007***	-0.006***	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Standardized Coef.	*-0.747*	*-0.747*	*-0.747*	*-0.749*	*-0.748*	*-0.710*	*-0.755*	*-0.785*	*-0.566*	*-0.478*	
Observations	1,296	1,296	1,296	1,296	1,296	1,296	1,296	1,296	1,296	1,296	
R-squared	0.768	0.768	0.768	0.768	0.767	0.766	0.768	0.769	0.772	0.772	

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Except in columns (2) and (3), mortality (Panel A) is the number of deaths divided by the total population; natality (Panel B) is the number of live births divided by the total population; birth cohort size (Panel C) is the birth cohort size reported by the 1939 census divided by the total 1939 population. In columns (2) and (3) mortality is the number of rural/urban deaths divided by the rural/urban population, natality is the number of rural/urban live births divided by the rural/urban population, birth cohort size is the rural/urban birth cohort divided by the rural/urban 1939 population (see column headings). Ukrainian population share is the share of self-reported ethnic Ukrainians in the rural population except in columns (4)–(7). In columns (6) and (7), the Ukrainian population share is the share of people in the total population whose mother tongue is Ukrainian. Famine is an indicator that equals one in 1932 and zero otherwise. Column (8) also controls for the share of people ages 10 and younger and the male/female ratio according to the 1926 census, each interacted with the famine indicator. Column (9) also controls for latitude × longitude × famine and all lower-order interactions. Column (10) controls for agricultural suitability of each of the 5 crops listed in the column heading interacted with the famine dummy variable. Column (11) examines the 1892 famine. In Panels A and B, all regressions control for urbanization, urbanization × famine, predicted grain, predicted grain × famine, and province and year fixed effects. In Panel C, all regressions control for 1926 urbanization × famine, predicted grain, predicted grain × famine, and province and year fixed effects. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Famine Mortality and Ethnic Ukrainian Population Share — Robustness to Controlling for Historical Factors

	Dependent Variable: Mortality in Year t+1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Historical Wealth							
	Baseline	Nominal Income 1897 × Famine	Real Income 1897 × Famine	Labor Productivity 1897 × Famine	Rural Labor Productivity 1897 (Lower Bound) × Famine	Rural Labor Productivity 1897 (Upper Bound) × Famine	Value of Agricultural Equipment 1910 × Famine	Livestock 1916 × Famine
Dep. var. mean.	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*
Ukrainians × Famine	0.051***	0.052***	0.051***	0.051***	0.050***	0.050***	0.046***	0.051***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.004)	(0.005)
Observations	337	337	337	337	337	337	337	337
R-squared	0.776	0.780	0.776	0.776	0.778	0.779	0.810	0.777
	B. Historical Cultural Norms, Institutions and Inequality							
	Baseline	Share of Catholics 1897 × Famine, Share of Orthodox Christians 1897 × Famine	Share of Serfs 1858 × Famine	Peasant Revolts 1895–1914 × Famine	Baseline with Info on Land 1905	Share of Peasant Land in Repartition Commune 1905 × Famine	Share of Peasant Households in Repartition Commune 1905 × Famine	Peasant and Private Land Gini 1905 × Famine
Dep. var. mean.	*0.022*	*0.022*	*0.022*	*0.022*	*0.021*	*0.022*	*0.022*	*0.022*
Ukrainians × Famine	0.051***	0.053***	0.052***	0.050***	0.042***	0.053***	0.056***	0.040***
	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.008)	(0.006)
Observations	337	337	337	337	286	286	286	286
R-squared	0.776	0.782	0.779	0.797	0.792	0.799	0.800	0.799

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. All estimates control for urbanization, urbanization × famine, predicted grain, predicted grain × famine, and province and year fixed effects. Additional controls are stated in the column headings. All income proxies are measured in per capita terms. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Famine Mortality and Ethnic Ukrainian Population Share — Robustness to Controlling for State Capacity

	Dependent Variable: Mortality in Year t+1			
	Baseline	1917 Bolshevik Vote Share × Famine	#Communists (averaged over 1922, 1927, and 1932) × Famine	#1930 Congress Delegates × Famine
	(1)	(2)	(3)	(4)
Dep. var. mean.	*0.022*	*0.022*	*0.022*	*0.022*
Ukrainians × Famine	0.051***	0.050***	0.051***	0.050***
	(0.006)	(0.006)	(0.005)	(0.006)
Standardized Coef.	*0.826*	*0.807*	*0.825*	*0.815*
Observations	337	337	337	337
R-squared	0.776	0.781	0.776	0.784

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Ukrainian population share is the share of self-reported ethnic Ukrainians in the rural population. Famine is an indicator that equals one in 1932 and zero otherwise. All regressions control for urbanization, urbanization × famine, predicted grain, predicted grain × famine, and province and year fixed effects. Additional controls are stated in the column headings. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Famine Mortality and Ethnic Ukrainian Population Share — District-level Estimates

	Dependent Variable: Mortality									
		I. Alternative Measures of Mortality		II. Alternative Measures of Ukrainian Population Share			III. Additional Robustness			
	Baseline with Province-Year FE	Urban Mortality	Rural Mortality	Total Ukrainians	Urban Ukrainians	Mother Tongue Ukrainian	Control for Gender Ratio	Control for Latitude × Longitude × Famine	Omit Ukraine	Omit Ukraine, Lower Volga, North Caucasus, West Siberia
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. var. mean.	*0.029*	*0.024*	*0.031*	*0.029*	*0.029*	*0.029*	*0.029*	*0.029*	*0.026*	*0.024*
Ukrainians × Famine	0.039***	0.008	0.029***	0.044***	0.043***	0.047***	0.040***	0.040***	0.022***	0.027***
	(0.010)	(0.006)	(0.007)	(0.011)	(0.008)	(0.010)	(0.010)	(0.009)	(0.007)	(0.010)
Std. Err. Clustered at District Level	[0.006]	[0.011]	[0.009]	[0.006]	[0.010]	[0.006]	[0.006]	[0.006]	[0.007]	[0.008]
Standardized Coef.	*0.509*	*0.154*	*0.354*	*0.547*	*0.438*	*0.595*	*0.509*	*0.512*	*0.192*	*0.234*
Urbanization	0.004	-0.005	0.001	-0.0001	-0.007	0.004	0.004	0.004	0.001	0.001
	(0.007)	(0.007)	(0.008)	(0.006)	(0.008)	(0.006)	(0.007)	(0.007)	(0.005)	(0.006)
Std. Err. Clustered at District Level	[0.008]	[0.009]	[0.012]	[0.008]	[0.009]	[0.008]	[0.008]	[0.008]	[0.007]	[0.007]
Urbanization × Famine	-0.008	0.009**	0.004	-0.004	-0.005	-0.007	-0.008	-0.007	0.002	0.004
	(0.006)	(0.004)	(0.003)	(0.004)	(0.006)	(0.005)	(0.006)	(0.006)	(0.003)	(0.004)
Std. Err. Clustered at District Level	[0.004]	[0.005]	[0.006]	[0.004]	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Observations	3,376	1,698	2,915	3,376	2,002	3,366	3,376	3,376	2,600	2,098
R-squared	0.765	0.898	0.899	0.769	0.786	0.772	0.765	0.769	0.730	0.684
Ukrainians										
Mean	0.248	0.248	0.248	0.237	0.223	0.229	0.248	0.248	0.069	0.035
Std. Dev.	0.370	0.370	0.370	0.355	0.293	0.364	0.370	0.370	0.168	0.122

Notes: The sample includes Ukraine and Russia. Observations are at the district and year level. Mortality is the number of deaths divided by total population unless otherwise stated in the column heading. Ukrainians is the share of ethnic Ukrainians in the rural population unless otherwise stated in the column heading. All regressions control for grain suitability × famine, and district and province-year fixed effects. Column (7) also controls for the male/female ratio × famine. Column (8) also controls for latitude × longitude × famine and all lower-order interactions. Column (9) omits Ukraine from the sample. Column (10) omits Ukraine and the provinces of Lower Volga, North Caucasus and West Siberia. Grain suitability is the FAO GAEZ wheat suitability index for low-input rain-fed agriculture. The standard errors in parentheses are adjusted for spatial correlation within 400 km. *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at the district level are presented in square brackets.

Table 7: Food Retention, Mortality

	Dependent Variable:			
	Procurement Share = Procurement/ Production	Retention = Production – Procurement	Mortality in Year t+1	
	(1)	(2)	(3)	(4)
Dep. var. mean.	*0.254*	*0.918*	*0.022*	*0.022*
Ukrainians × Famine	0.185***	-1.085***		
	(0.018)	(0.191)		
Retention			-0.030**	-0.029***
			(0.012)	(0.010)
Retention^2			0.009**	0.009**
			(0.004)	(0.004)
Retention × Famine				-0.015
				(0.027)
Retention^2 × Famine				0.004
				(0.009)
Observations	107	107	107	107
R-squared	0.936	0.782	0.634	0.643

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Production is revised by the authors based on archival documents (see text). Retention is measured in kilograms per person per day. Mortality is the number of deaths divided by the total population. All columns control for urbanization, urbanization × famine, and province and year fixed effects. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Grain Production, Procurement and Retention Targets as Reported in the First Five Year Plan (published in 1928)

	Dependent Variables: Centrally Planned Targets (kg/person/day)			
	Production	Procurement	Procurement Share = Procurement/ Production	Retention = Production – Procurement
	(1)	(2)	(3)	(4)
Dep. var. mean.	*1.673*	*0.268*	*0.089*	*1.404*
Ukrainians	0.776***	0.275***	0.165***	-0.275***
	(0.191)	(0.071)	(0.020)	(0.071)
Official Grain Production 1928	0.907***			
	(0.052)			
Production Target		0.540***	0.285***	0.460***
		(0.035)	(0.016)	(0.035)
Observations	90	90	90	90
R-squared	0.691	0.814	0.844	0.723

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level for the years of the First Five-Year Plan, 1928-33. Ukrainians is the 1926 share of ethnic Ukrainians in the rural population. Official grain production 1928 is measured in kg/person/day. All estimates control for year fixed effects. Huber-White robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

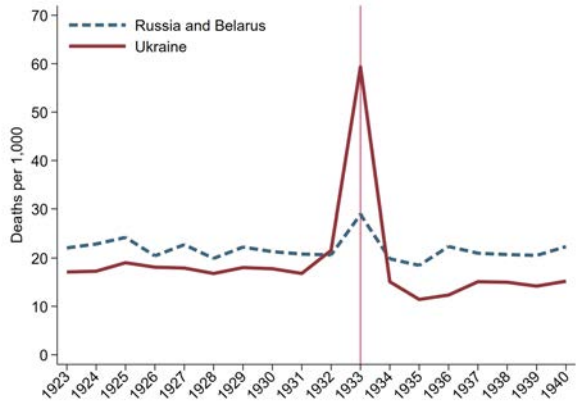
Table 9: Heterogeneous Effects of Grain Productivity on Famine Mortality in Ethnic Ukrainian Areas

	Dependent Variable: Mortality in Year t+1			
	Province- Level			District-Level
	(1)	(2)		(3)
Ukrainians × Official Grain Production 1928 × Famine	0.300***	0.263***	Ukrainians × Grain Suitability × Famine	0.071***
	(0.051)	(0.081)		(0.022)
Ukrainians × Famine	-0.174***	-0.065	Ukrainians × Famine	-0.011
	(0.035)	(0.156)		(0.017)
Official Grain Production 1928 × Famine	0.0002	0.001	Grain Suitability × Famine	0.004
	(0.002)	(0.002)		(0.005)
State Capacity × Famine		-0.001		
		(0.001)		
Ukrainians × State Capacity × Famine		0.022		
		(0.027)		
Observations	337	337	Observations	3,376
R-squared	0.845	0.846	R-squared	0.783

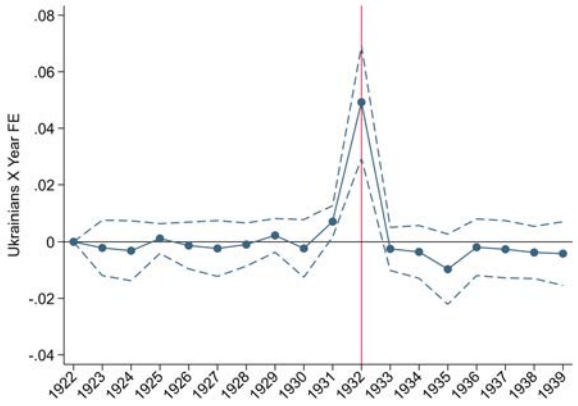
Notes: The province sample includes Ukraine, Russia and Belarus; the distict sample includes Ukraine and Russia. In columns (1)-(2), the observations are at the province and year level. Mortality is the number of deaths divided by the total population. State Capacity is the first principal component of the three variables measuring political loyalty and state capacity: the share of votes for the Bolshevik Party in the 1917 Constituency Assembly election, the number of Communist Party Members (averaged over 1922, 1927 and 1931) per 1,000 individuals in each province, and the number of Party secretaries (at the province, district, city and, if the city was large, the borough level) who attended the 1930 Party Congress to vote formally for the policy of comprehensive collectivization. The regressions control for urbanization, urbanization × famine, Ukrainians × urbanization × famine, predicted grain, predicted grain × famine, Ukrainians × predicted grain × famine; and province and year fixed effects. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. In column (3), observations are at the district and year level. Instead of official grain production in 1928, the triple interaction is with the suitability for grain cultivation. The regression controls for urbanization, urbanization × famine, Ukrainians × urbanization × famine, and district and province-year fixed effects. The standard errors in parentheses are adjusted for spatial correlation within 400 km. *** p<0.01, ** p<0.05, * p<0.1

Figure 1: Famine Severity over Time – Means, Estimates of $Ukrainian \times Year\ FEs$

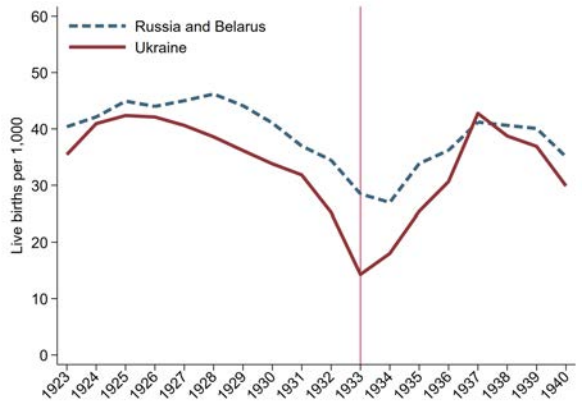
(a) Average Mortality



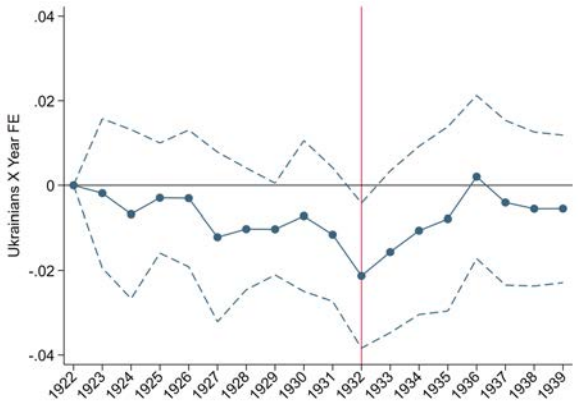
(b) Dynamic Effects on Mortality



(c) Average Natality

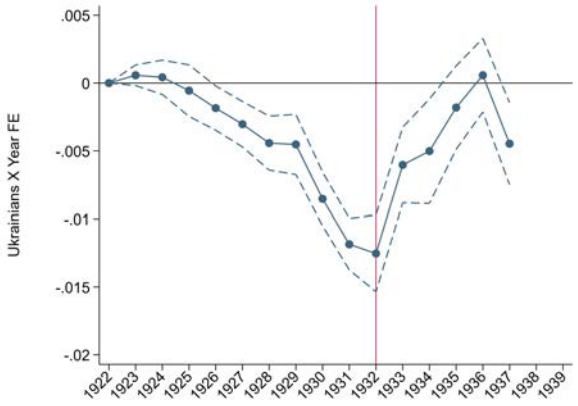


(d) Dynamic Effects on Natality



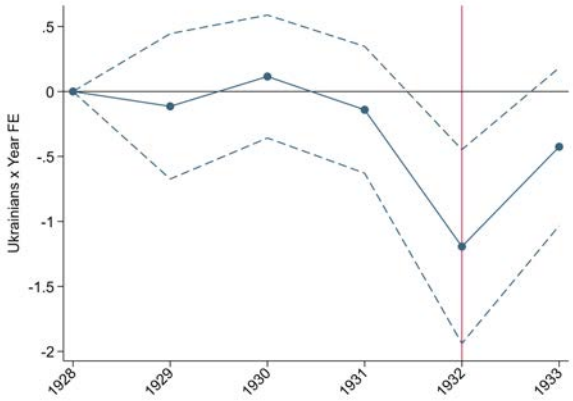
(e) Average Birth Cohort Size

(f) Dynamic Effects on Birth Cohort Size



Notes: Mortality is the number of deaths per 1,000 individuals. Natality is the number of live births per 1,000 individuals. Birth cohort size is the number of individuals born each year divided by the total population as reported by the 1939 census. The dynamic effects figures plot estimates of the interaction coefficients of Ukrainian population share and year dummy variables with their 95% confidence intervals. The estimates and their standard errors are presented in Appendix Table A.4.

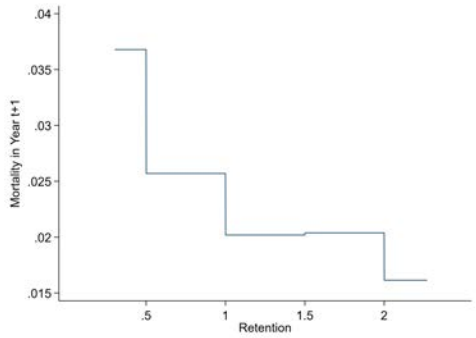
Figure 2: Dynamic Effect on Grain Retention



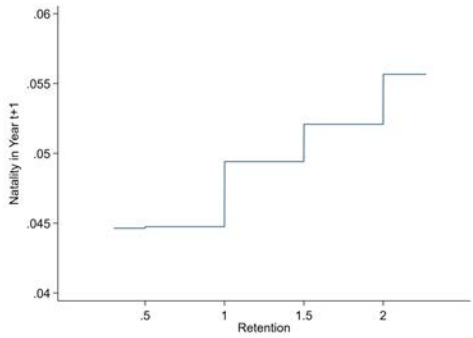
Notes: The figure plots estimates of the interaction coefficients of Ukrainian population share and year dummy variables with their 95% confidence intervals. The estimates and their standard errors are presented in Appendix Table A.6 column (7).

Figure 3: Piece-wise Linear Estimation of Famine Severity and Grain Retention

(a) Mortality



(b) Natality

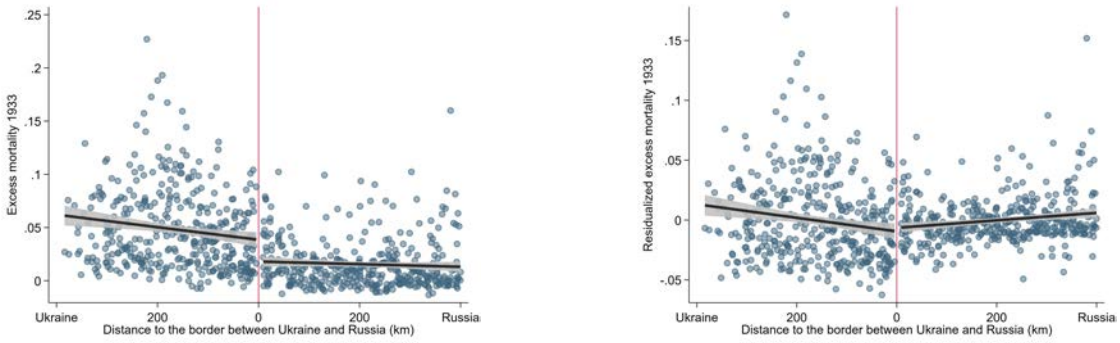


Notes: The figures plot the coefficients for different levels of grain retention (production - procurement), controlling for province and year fixed effects, from the piece-wise linear function in equation (2). The coefficients and standard errors are shown in Appendix Table A.6.

Figure 4: District-Level Excess Mortality in 1933 and Distance from the Ukrainian-Russian Border

(a) Excess Mortality in 1933

(b) Excess Mortality 1933 (control for Ukrainian Pop Share)

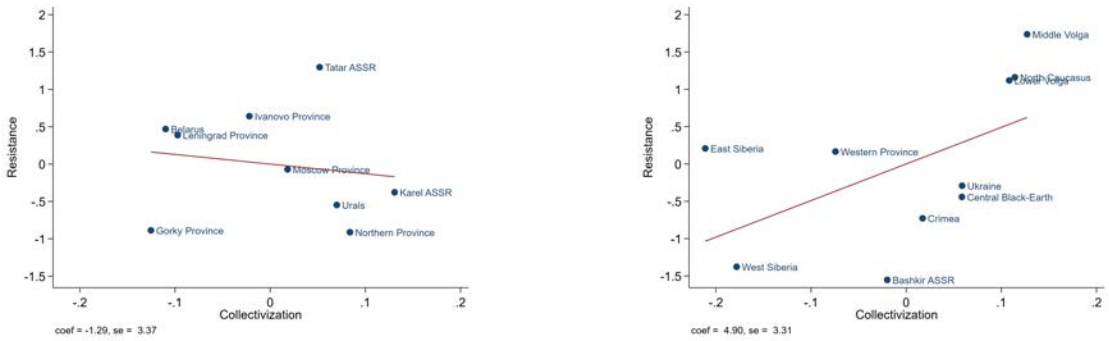


Notes: The figures plot excess mortality against distance to the Ukraine-Russian border and the fitted lines and their 95% confidence interval for each district in Ukraine and Russia. Excess mortality is the difference between 1933 and 1928 mortality for each district. In Figure 4b, excess mortality is demeaned by the share of ethnic Ukrainians in each district.

Figure 5: Collectivization and Peasant Resistance (De-Classified Police Reports)

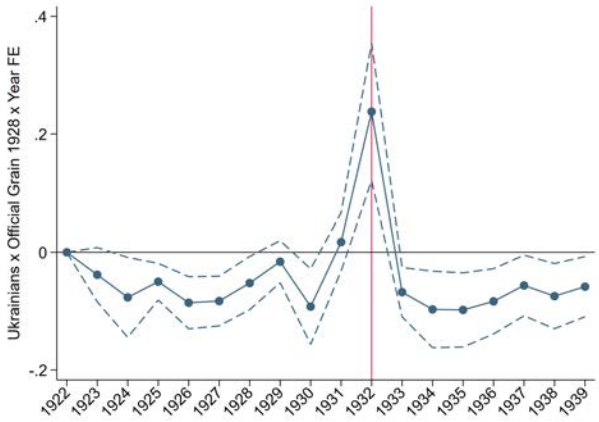
(a) Low Ukrainian Share

(b) High Ukrainian Share



Notes: Figures 5a and 5b plot the correlation between collectivization and peasant resistance conditional on official 1928 per capita grain production and urbanization for the provinces with low Ukrainian share ($<$ sample median) and high Ukrainian share ($\geq$ sample median). The p-value for the statistical difference of the two slopes is 0.092.

Figure 6: Heterogeneous Effects

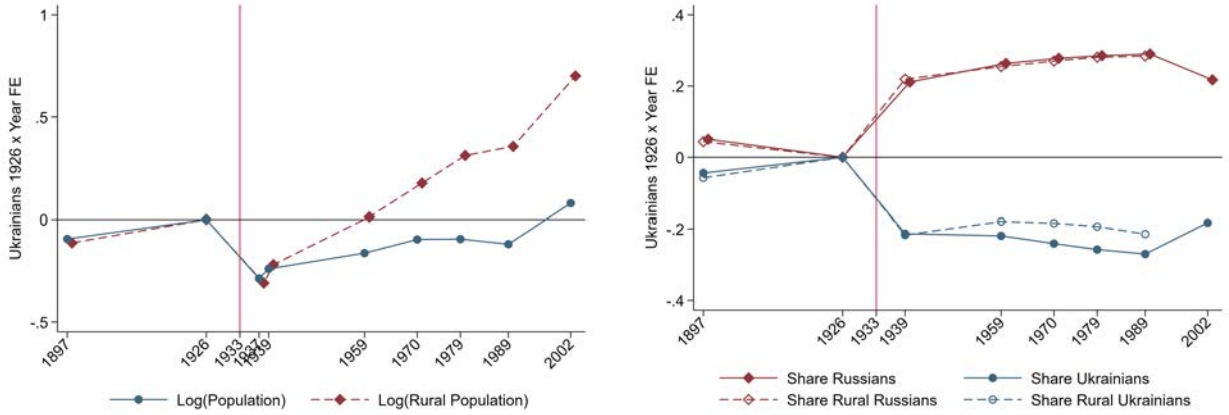


Notes: Figure 6 plots the triple interaction coefficients of per capita grain production in 1928, Ukrainian population share and year dummy variables with their 95% confidence intervals. The estimates and their standard errors are presented in Appendix Table A.4.

Figure 7: Population (Census Data) and Ukrainian Population Share

(a) Population

(b) Ethnic Composition



Notes: Figures 7a and 7b plot the coefficients for the interaction of 1926 Ukrainian population share and census year dummy variables, controlling for province and year fixed effects. The dependent variables are stated in the table legends. The estimates use population census data from the years stated on the x-axis. 1926 is omitted for comparison. Appendix Table A.7 reports the coefficients and their standard errors.

Online Appendix (Not for Publication)

A Revising Grain Production

Appendix Table A.2 presents the aggregate food counting exercise for Russia and the entire Soviet Union that are analogous to what is shown for Ukraine in Table 1. Panel II presents data for Russia. Panel III for all of the Soviet Union. We also compare our production corrections to those from Davies and Wheatcroft (2009), which are only available for the aggregate level and only as a range.[2]

For brevity, we will focus our discussion about the different corrections on Panel II. We approximate "true" production by using grain procurement ratios (as a share of production) reported by the State Planning Committee (Gosplan), and procurement levels, which as we discuss in the main text are widely accepted as accurate since they were only ever used internally. Procurement stocks were directly observed and counted by the government.

Similarly, the accuracy of procurement ratios reported in the Gosplan has not been disputed. The Gosplan report was a candid evaluation of the First Five-Year Plan shown only to the highest ranking Soviet officials. The report was classified until after the fall of the Soviet Union. Procurement ratios are reported for each province and year 1928-1932 in RSAE 4372/30/871 (p. 25). Dividing procurement levels by procurement ratios recovers production levels.

The main advantages of our correction are that it can be calculated at regional levels and require few, straightforward assumptions. Our correction assumes that a decline in procurement reflects a decline in production rather than a decline the ability of the government to collect. If, for example, peasants hid more (less) grain to evade procurement in 1932 than in other years, then our estimates would understate (overstate) true production. Given historical accounts that intense campaigns to prevent peasants from hiding grain started with the the introduction of collectivization several years before the famine, the *prima facie* assumption is that there was little evasion (and little change in evasion) between 1930 and 1932. In other analyses of the paper, we address the measurement error in production figures by not using reported production during this period.

The Gosplan report does not cover 1933. Thus, to estimate production for 1933, we follow the spirit of Davies and Wheatcroft (2009) and calculate 1933 production as collectivization rate $\times$grain sown area $\times$ kolkhoz yield + (1 - collectivization rate) $\times$ grain sown area $\times$ official yield. 1933 kolkhoz yields are reported in the Gosplan report on the state of collective farms in 1932–33, RSAE 1562/77/70 (p. 31). Collectivization rate is the share of collectivized rural households. This calculation assumes that sown area was similar for collectivized and non-collectivized households. In reality, collectivized households were given more land to farm. It also assumes that individual farm yields were equal to official yields.

Panel II rows (6) and (8)-(9) show that, except for 1930, our corrections fall within the DW range for years for which both are available. It is also reassuring that our corrected production is below the officially reported production in row (7) for 1931-33, since this supports the belief that the official aggregate numbers

[2]The derivation of these estimates, which are presented in Davies and Wheatcroft (2009) Table 1, p. 448-449, is unclear. The authors explain that "Our estimates are based on a range of different data that were accepted internally by the best experts of the time, and our own assessments of the reliability of these different data" (p. 446). See their earlier work for more discussion.

were inflated only in the years when production fell relative to expectations.

Note that after 1933, the government switched from reporting realized production to the maximum possible yields given weather, geographic conditions and inputs such as labor.[3] This change does not affect our study since we do not use reported production in our analysis.

B Predicting Grain Production

To estimate the grain production function, we use data from 1901–15. We regress log grain production (total harvests) on log province area, log FAO GAEZ grain suitability index, their interaction, temperature and precipitation for each of the four seasons, their pairwise interactions and square terms (without a constant). The seasons are: fall (October, November, and December of the previous calendar year), winter (January, February, March), spring (April, May, June), summer (July, August, September). Appendix Table A.3 presents the estimated grain production function. We then use this production function to predict grain harvest from 1922 to 1940. The predicted grain and actual grain are closely correlated, with two exceptions: Karelia and East Siberia provinces. The in-sample R-squared is 0.90. The out-of-sample R-squared is 0.77, see Appendix Figure A.2. The high out-of-sample predictive power is consistent with the lack of major technological changes in Soviet agriculture before the 1930s (e.g., Allen, 2003).

C Weather

Appendix Table A.5 Panel A controls for weather directly instead of grain production predicted by weather (and other natural conditions). Column (1) presents the baseline without controlling for predicted grain and its interaction. Column (2) controls for spring and summer temperature and precipitation, since weather in the spring and summer of 1931 and 1932 is discussed most often as a cause of poor harvests during the famine (e.g., Davies and Wheatcroft, 2009; Tauger, 1991). It also controls for the following year's winter temperature and precipitation since 1933 winter weather conditions may have directly affected mortality. Column (3) controls for monthly temperature and precipitation and their squared terms (48 additional controls). Column (4) controls for monthly temperature and precipitation and their interactions (36 additional controls). Column (5) controls for monthly weather shock indicators following the standard in the literature, where the weather shock indicator is equal to one if the month's temperature or precipitation is more than one standard deviation away from the long-term (1900–50) mean (12 additional controls). Column (6) controls for the deviations from the long-term median of monthly temperature and precipitation fro the 12 months of year t (24 additional controls). Finally, as in Rozenas and Zhukov (2019), column (7) controls the monthly deviations of the two variables for years $t-1$ and t (48 additional controls). The main interaction coefficient of the rural share of ethnic Ukrainians and the famine indicator changes little with these controls.

[3] Davies and Wheatcroft (2009) discusses this change from "the barn harvest" to "the harvest on the root".

D *Dekulakization*

Appendix Table A.5 Panel B shows that our main result is robust to different ways of controlling for the extent of *dekulakization* in the region. These measures are the number of exiled kulak households during 1930–31 (Davies and Wheatcroft, 2009, Table 28), the number of exiled kulak households during 1930–31 according to a secret police report in (Berelowitch and Danilov, eds, 2000-2012, Document 253), *ex ante* 1930 quotas for kulak exile, secret police estimates of total number of kulaks in countryside, and the number of arrested peasants.

Table A.1: Ethnic Composition in the USSR

	1926 census						1939 census	
	Total		Urban		Rural		Total	
	mil.	%	mil.	%	mil.	%	mil.	%
	A. All USSR							
Russians	77.8	53.1	16.6	63.5	61.2	50.8	99.6	58.4
Ukrainians	31.2	21.3	3.3	12.6	27.9	23.2	28.1	16.5
Belorussians	4.7	3.2	0.5	1.9	4.2	3.5	5.3	3.1
	B. USSR (Regression Sample, Subset of Panel A)							
Russians	77.1	57.2	16.1	67.9	61.1	54.9	94.8	65.0
Ukrainians	31.1	23.1	3.2	13.7	27.9	25.1	27.1	18.6
Belorussians	4.7	3.5	0.5	2.0	4.2	3.8	5.2	3.6
	C. "Grain-producing" Provinces (Subset of Panel B)							
Ukrainians	28.5	43.8	3.0	29.2	25.5	46.6	25.3	37.1
Russians	27.3	41.9	5.2	50.0	22.0	40.3	32.8	48.1
Tatars	2.2	3.4	0.2	2.0	2.0	3.6	2.9	4.3

Notes: These data are reported by the 1926 and 1939 Population Censuses. Panel C includes Bashkir ASSR, Central Black-Earth region, Crimea, Lower Volga, Middle Volga, North Caucasus, Tatar ASSR, and Ukraine.

Table A.2: Caloric Accounting

		1927	1928	1929	1930	1931	1932	1933	1937	1939
		I. Russia								
(1)	Production (mln tons)	.	60.5	48.2	54.8	43.0	44.3	50.8	.	.
(2)	Procurement (mln tons)	6.2	9.6	10.2	13.9	14.8	13.7	15.3	.	.
(3)	Rural retention (kg/person/day)	.	1.63	1.19	1.27	0.87	0.96	1.12	.	.
(4)	Retention if no procurement	.	1.94	1.51	1.70	1.33	1.39	1.61	.	.
(5)	Food needs for heavy labor (kg/person/day)	0.78	0.78	0.78	0.78	0.78	0.78	0.79	0.75	0.75
		II. USSR								
(6)	Production (mln tons)	.	73.4	71.8	83.3	64.5	57.1	75.1	.	.
(7)	Official	74.1	73.3	71.7	83.5	69.5	69.9	89.8	120.3	100.9
(8)	DW Min	.	.	.	73.0	57.0	55.0	70.0	.	.
(9)	DW Max	.	.	.	77.0	65.0	60.0	77.0	.	.
(10)	Procurement (mln tons)	11.1	10.8	16.1	22.2	22.8	19.0	23.7	31.9	30.9
(11)	Rural retention (kg/person/day)	.	1.38	1.21	1.31	0.89	0.82	1.11	.	.
(12)	Retention if no procurement	.	1.62	1.55	1.78	1.38	1.23	1.62	.	.
(13)	Food needs for heavy labor (kg/person/day)	0.78	0.78	0.78	0.78	0.78	0.77	0.77	0.76	0.76

Notes: Data for population, production and procurement are official statistics. Rows (1) and (6) report production revised by the authors using archival data (see text). Row (7) reports official production figures. Rows (8)-(9) report production revised by Davis and Wheatcroft (2004) (see their Table 1). Retention is the difference between our revised production and procurement. Food needs are calculated by the authors and take into account the demographic composition (e.g., age, gender, rural/urban) as reported in the population censuses. They are based on official guidelines for maximum caloric needs for each group as reported by Lositskij (1926, 1928). Panel I includes Russia. Panel II includes all USSR.

Table A.3: The Effect of Weather and Natural Conditions on Grain Production

	Dependent Variable: Log Grain Production		
	(1)		(2)
Log area	0.352***	Fall temperature × Fall precipitation	0.0005*
	(0.067)		(0.0002)
Log grain suitability	-4.643***	Winter temperature × Winter precipitation	0.001*
	(0.640)		(0.0003)
Log area × Log grain suitability	0.278***	Spring temperature × Spring precipitation	0.0004
	(0.023)		(0.0003)
Fall temperature	0.015	Summer temperature × Summer precipitation	0.001***
	(0.037)		(0.0003)
Winter temperature	0.027	Fall temperature^2	0.004**
	(0.043)		(0.002)
Spring temperature	-0.169**	Winter temperature^2	0.0003
	(0.079)		(0.002)
Summer temperature	-0.978***	Spring temperature^2	-0.001
	(0.194)		(0.003)
Fall precipitation	-0.006	Summer temperature^2	0.028***
	(0.006)		(0.005)
Winter precipitation	-0.005	Fall precipitation^2	0.00002
	(0.007)		(0.00002)
Spring precipitation	0.01	Winter precipitation^2	0.00004
	(0.007)		(0.00003)
Summer precipitation	-0.025***	Spring precipitation^2	-0.00003**
	(0.009)		(0.00002)
		Summer precipitation^2	0.00003**
			(0.00001)
Observations		220	
R-squared		0.998	

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Log grain is the logarithm of the grain harvest. Log area is the logarithm of province area. Log grain suitability is the logarithm of the province's FAO GAEZ wheat suitability index for rain-fed low-input agriculture. Fall is October, November, December of the previous calendar year; Winter is January, February, March; Spring is April, May, June; Summer is July, August, September. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: The Dynamic Relationship between Ukrainian Population Share, Famine Intensity and Grain Retention

	Dependent Variable:							
	Mortality in Year t+1	Rural Mortality in Year t+1	Natality in Year t+1	Rural Natality in Year t+1	Total Birth Cohort Size in Year t+1/Total 1939 Population	Rural Birth Cohort Size in Year t+1/Rural 1939 Population		Mortality in Year t+1
	(1)	(2)	(3)	(4)	(5)	(6)		(7)
Ukrainians × 1923	-0.002		-0.002		0.001	0.001*	Ukrainians × Official Grain 1928 × 1923	-0.038
	(0.003)		(0.007)		(0.0004)	(0.0005)		(0.024)
Ukrainians × 1924	-0.003		-0.007		0.0004	0.001	Ukrainians × Official Grain 1928 × 1924	-0.077**
	(0.003)		(0.007)		(0.001)	(0.001)		(0.035)
Ukrainians × 1925	0.001		-0.003		-0.001	-0.0002	Ukrainians × Official Grain 1928 × 1925	-0.050***
	(0.002)		(0.005)		(0.001)	(0.001)		(0.016)
Ukrainians × 1926	-0.001	-0.003	-0.003	-0.0002	-0.002**	-0.002*	Ukrainians × Official Grain 1928 × 1926	-0.086***
	(0.003)	(0.002)	(0.006)	(0.002)	(0.001)	(0.001)		(0.023)
Ukrainians × 1927	-0.002	-0.004	-0.012	-0.009**	-0.003***	-0.003***	Ukrainians × Official Grain 1928 × 1927	-0.083***
	(0.004)	(0.003)	(0.008)	(0.004)	(0.001)	(0.001)		(0.022)
Ukrainians × 1928	-0.001	-0.003**	-0.010*	-0.007***	-0.004***	-0.005***	Ukrainians × Official Grain 1928 × 1928	-0.052**
	(0.003)	(0.001)	(0.005)	(0.002)	(0.001)	(0.001)		(0.023)
Ukrainians × 1929	0.002	0.001	-0.010***	-0.007***	-0.005***	-0.005***	Ukrainians × Official Grain 1928 × 1929	-0.016
	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)		(0.018)
Ukrainians × 1930	-0.002	-0.003	-0.007	-0.003	-0.009***	-0.011***	Ukrainians × Official Grain 1928 × 1930	-0.092***
	(0.004)	(0.003)	(0.007)	(0.003)	(0.001)	(0.001)		(0.033)
Ukrainians × 1931	0.007***	0.008***	-0.012*	-0.007***	-0.012***	-0.015***	Ukrainians × Official Grain 1928 × 1931	0.017
	(0.002)	(0.001)	(0.006)	(0.002)	(0.001)	(0.001)		(0.025)
Ukrainians × 1932	0.049***	0.059***	-0.021***	-0.017***	-0.013***	-0.016***	Ukrainians × Official Grain 1928 × 1932	0.238***
	(0.007)	(0.006)	(0.006)	(0.003)	(0.001)	(0.002)		(0.059)
Ukrainians × 1933	-0.003	-0.005***	-0.016**	-0.012***	-0.006***	-0.008***	Ukrainians × Official Grain 1928 × 1933	-0.068***
	(0.003)	(0.002)	(0.007)	(0.004)	(0.001)	(0.002)		(0.021)
Ukrainians × 1934	-0.004	-0.007***	-0.011	-0.007**	-0.005**	-0.007***	Ukrainians × Official Grain 1928 × 1934	-0.097***
	(0.003)	(0.002)	(0.008)	(0.003)	(0.002)	(0.002)		(0.033)
Ukrainians × 1935	-0.010**	-0.013***	-0.008	-0.005	-0.002	-0.003	Ukrainians × Official Grain 1928 × 1935	-0.098***
	(0.005)	(0.004)	(0.009)	(0.004)	(0.002)	(0.002)		(0.032)
Ukrainians × 1936	-0.002	-0.006**	0.002	0.003	0.001	0.001	Ukrainians × Official Grain 1928 × 1936	-0.083***
	(0.003)	(0.002)	(0.007)	(0.003)	(0.001)	(0.001)		(0.028)
Ukrainians × 1937	-0.003	-0.006***	-0.004	-0.002	-0.004***	-0.005***	Ukrainians × Official Grain 1928 × 1937	-0.056**
	(0.003)	(0.002)	(0.008)	(0.003)	(0.002)	(0.002)		(0.026)
Ukrainians × 1938	-0.004	-0.006***	-0.006	-0.003			Ukrainians × Official Grain 1928 × 1938	-0.074***
	(0.003)	(0.002)	(0.007)	(0.003)				(0.028)
Ukrainians × 1939	-0.004	-0.007**	-0.006	-0.002			Ukrainians × Official Grain 1928 × 1939	-0.058**
	(0.004)	(0.003)	(0.007)	(0.003)				(0.026)
Observations	337	285	337	285	1,296	1,296		337
R-squared	0.811	0.819	0.870	0.926	0.822	0.767		0.914

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Mortality is the number of deaths divided by the total population. Natality is the number of live births divided by the total population. Birth cohort size is the total/rural birth cohort size reported by the 1939 census divided by the total/rural 1939 population. Columns (1)-(4) control for urbanization interacted with year indicators, predicted grain interacted with year indicators, and province and year fixed effects. Columns (5)-(6) control for 1926 urbanization interacted with year indicators, predicted grain interacted with year indicators, and province and year fixed effects. Column (7) controls for urbanization × Ukrainians × year indicators, predicted grain × Ukrainians × year indicators and all lower-order interaction terms; and province and year fixed effects. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Famine Mortality and Ukrainian Population Share – Robustness to Controlling for Weather, *Dekulakization,* Demographic Structure

	Dependent Variable: Mortality in year t+1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	A. Weather						
	Baseline[+]	Spring, Summer, Winter Rain & Temp	Monthly Temp, Rain, and Quadratics	Monthly Temp, Rain, Temp × Rain	Monthly Weather Shock	Monthly Temp and Rain Deviations from Historical Monthly Median in Year t	Monthly Temp and Rain Deviations for Year t-1, t
Dep. var. mean.	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*
Ukrainians × Famine	0.051***	0.051***	0.051***	0.048***	0.050***	0.048***	0.047***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)
Observations	337	337	337	337	337	337	337
R-squared	0.776	0.782	0.828	0.814	0.784	0.806	0.825
	B. De-Kulakization						
	Baseline	Exiled Kulaks (DW) × Famine	Exiled Kulaks (OGPU) × Famine	Planned Kulaks 1930 × Famine	Total Kulaks (OGPU Estimate) × Famine	Arrested Kulaks 1930 × Famine	First Principal Component of Kulak Variables × Famine
Dep. var. mean.	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*
Ukrainians × Famine	0.051***	0.054***	0.053***	0.044***	0.058***	0.051***	0.053***
	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)	(0.005)
Observations	337	337	337	337	267	302	249
R-squared	0.776	0.792	0.789	0.807	0.780	0.787	0.800
	C. Demographic Structure						
	Baseline	Share of Infants × Famine, Gender Ratio × Famine	Share of Children 5 and Younger × Famine, Gender Ratio × Famine	Share of Children 10 and Younger × Famine, Gender Ratio × Famine	Share of Adults 50 and Older × Famine, Gender Ratio × Famine	Share of Adults 70 and Older × Famine, Gender Ratio × Famine	
Dep. var. mean.	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	*0.022*	
Ukrainians × Famine	0.051***	0.055***	0.046***	0.048***	0.051***	0.051***	
	(0.006)	(0.006)	(0.007)	(0.007)	(0.005)	(0.005)	
Observations	337	337	337	337	337	337	
R-squared	0.776	0.780	0.785	0.783	0.784	0.784	

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. [+]In Panel A, all regressions control for urbanization, urbanization × famine and province and year fixed effects. In Panels and C, all regressions control for urbanization, urbanization × famine, predicted grain, predicted grain × famine, and province and year fixed effects. Additional controls are stated in the column headings. In Panel A col. (5), a shock is a dummy variable which equals1 if temp or rain is one std dev or more different from historical province mean. In cols. (6) and (7), we control for the deviation of each month from the historical sample median. In col. (6), we control for the 12 months of the current year. In col. (7), we additionally control for the 12 months in the previous year. See the Appendix discussion. In Panel B col. (2), exiled kulaks (DW) are the number of dekulakized and exiled households in 1930–31 per 1930 population according to Davies and Wheatcroft (2004), Table 28. In col. (3), exiled kulaks (OGPU) are the number of dekulakized and exiled households in 1930–31 according to an OGPU report (Soviet Countryside from the Perspective of VChK-OGPU-NKVD, Document 253) per 1930 population. In col. (4), planned kulaks are the planned number of dekulakizations per capita in February 1930 (the average between lower and upper bounds). In col. (5), total kulaks (OGPU estimate) are the total number of kulaks in the rural population according to the OGPU estimate (Soviet Countryside from the Perspective of VChK-OGPU-NKVD, Document 253). In col. (6), arrested kulaks are the number of peasants processed by "troiki" per capita in 1930 according to the OGPU estimate (Soviet Countryside from the Perspective of VChK-OGPU-NKVD, Document 279). The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Mortality, Natality and Grain Retention

	Dependent Variable:								
	Mortality in Year t+1			Natality in Year t+1				Retention	
	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)
Dep. var. mean.	*0.022*	*0.022*	*0.022*	*0.038*	*0.038*	*0.038*	*Dep. var. mean.*	*1.019*	*1.019*	*1.019*
Retention: < 0.5	0.037***	0.036***	0.036***	0.045***	0.046***	0.046***	Ukrainians × 1929	-0.114	-0.111	-0.115
	(0.009)	(0.009)	(0.009)	(0.003)	(0.003)	(0.003)		(0.285)	(0.284)	(0.286)
Retention: (0.5, 1.0]	0.026***	0.025***	0.025***	0.045***	0.045***	0.046***	Ukrainians × 1930	0.115	0.119	0.116
	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)		(0.241)	(0.240)	(0.242)
Retention: (1.0, 1.5]	0.020***	0.019***	0.020***	0.049***	0.051***	0.051***	Ukrainians × 1931	-0.141	-0.135	-0.137
	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)		(0.249)	(0.247)	(0.249)
Retention: (1.5, 2.0]	0.020***	0.020***	0.021***	0.052***	0.053***	0.053***	Ukrainians × 1932	-1.193***	-1.192***	-1.160***
	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)		(0.380)	(0.377)	(0.352)
Retention: > 2.0	0.016***	0.015***	0.015***	0.056***	0.058***	0.058***	Ukrainians × 1933	-0.425	-0.416	-0.412
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)		(0.310)	(0.306)	(0.306)
Controls:							Controls:			
Urbanization × Famine		Y	Y		Y	Y	Urbanization × Famine		Y	Y
Official Grain 1928 × Famine			Y			Y	Official Grain 1928 × Famine			Y
Observations	107	107	107	107	107	107	Observations	107	107	107
R-squared	0.964	0.964	0.968	0.995	0.995	0.995	R-squared	0.788	0.790	0.793

Notes: The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. Mortality is the number of deaths divided by the total population. Natality is the number of live births divided by the total population. Retention is the author's revised grain production (see text) minus grain procurement, measured in kilograms per person per day. Ukrainians is the 1926 Ukrainian population share. All regressions control for urbanization, province and year fixed effects. In columns (1)-(6), Huber-White robust standard errors are in parentheses; in columns (7)-(9), the standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Population and Ethnic Composition in the Long-Run

	Dependent Variable					
	Log Population (1)	Log Rural Population (2)	Share Ukrainians (3)	Share Russians (4)	Share Rural Ukrainians (5)	Share Rural Russians (6)
Dep. var. mean.	*15.496*	*14.754*	*0.078*	*0.711*	*0.086*	*0.663*
Ukrainians × 1897	-0.094	-0.113	-0.044***	0.051***	-0.058***	0.044***
	(0.098)	(0.099)	(0.007)	(0.015)	(0.009)	(0.016)
Ukrainians × 1937	-0.287***	-0.309***				
	(0.093)	(0.071)				
Ukrainians × 1939	-0.240***	-0.220***	-0.213*	0.211	-0.217	0.219
	(0.075)	(0.039)	(0.129)	(0.130)	(0.150)	(0.150)
Ukrainians × 1959	-0.163	0.013	-0.219*	0.264*	-0.179	0.255
	(0.115)	(0.093)	(0.128)	(0.136)	(0.155)	(0.177)
Ukrainians × 1970	-0.097	0.177	-0.241*	0.278**	-0.184	0.270
	(0.115)	(0.162)	(0.123)	(0.127)	(0.154)	(0.176)
Ukrainians × 1979	-0.095	0.312	-0.257**	0.285**	-0.193	0.281
	(0.120)	(0.207)	(0.121)	(0.120)	(0.151)	(0.173)
Ukrainians × 1989	-0.120	0.356	-0.270**	0.290***	-0.214	0.284*
	(0.134)	(0.260)	(0.118)	(0.111)	(0.148)	(0.160)
Ukrainians × 2002	0.079	0.701***	-0.183	0.217*		
	(0.121)	(0.266)	(0.140)	(0.122)		
Observations	171	171	152	152	133	133
R-squared	0.958	0.952	0.953	0.969	0.941	0.957

Notes: The dependent variables are reported by population censuses. Ukrainians is the share of ethnic Ukrainians in the rural population in 1926. The sample includes Ukraine, Russia and Belarus. Observations are at the province and year level. All regressions control for province and year fixed effects. The standard errors in parentheses are adjusted for spatial correlation within 1,500 km. *** p<0.01, ** p<0.05, * p<0.1

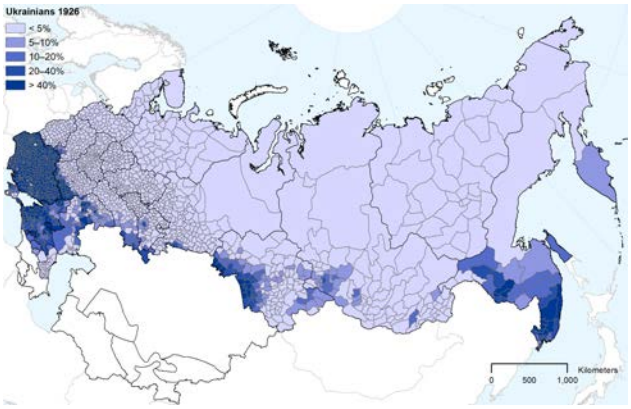
Figure A.1: Maps

(a) Province Excess Mortality 1933 and Grain-Producing Regions

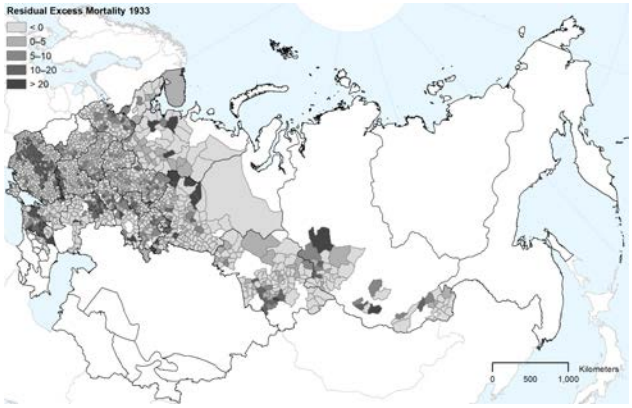
(b) Province Ethnic Ukrainians (1926) and Grain-Producing Regions

(c) District Excess Mortality 1933

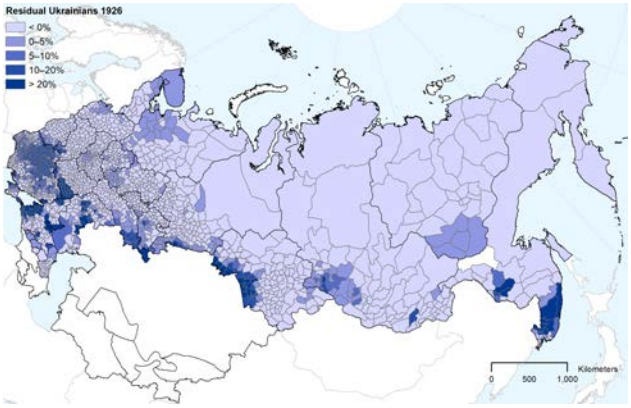
(d) District Ethnic Ukrainians 1926



(e) District Excess Mortality 1933 Demeaned by Province Fixed Effects



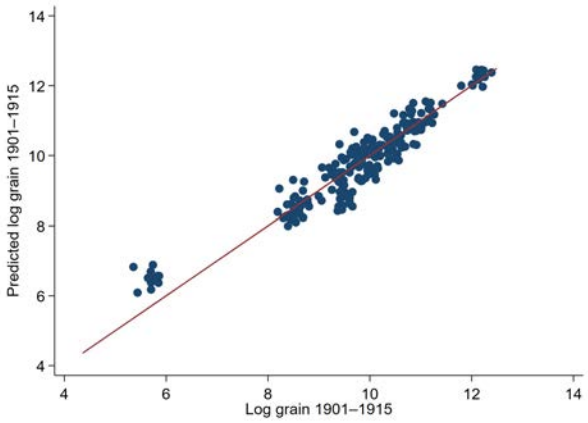
(f) District Ethnic Ukrainians 1926 Demeaned by Province Fixed Effects



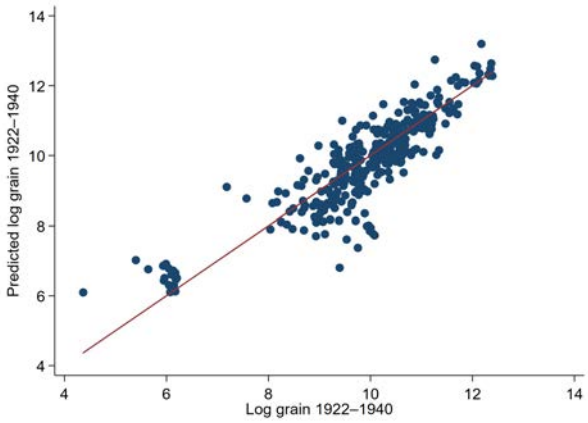
Notes: Excess mortality 1933 is mortality in 1933 minus mortality in 1928. Ethnic Ukrainians 1926 is the share of ethnic Ukrainians in the rural population according to the 1926 Population Census.

Figure A.2: Predicted Grain

(a) In-Sample Fit, $R^2 = 0.90$



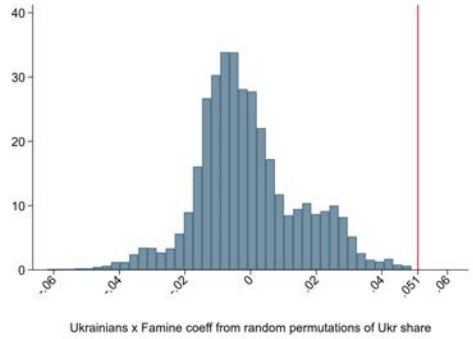
(b) Out-of-Sample Fit, $R^2 = 0.77$



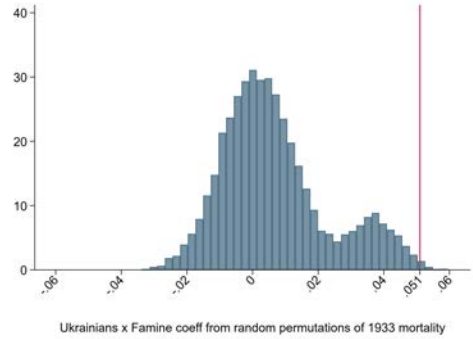
Notes: Log reported grain is plotted against log predicted grain with a 45-degree line for 1901–1915, a sample on which grain production function is estimated (in-sample fit) in Figure A.2a, and for 1922–1940 (out-of-sample fit) in Figure A.2b. See Appendix section B for details.

Figure A.3: Random Permutation Test

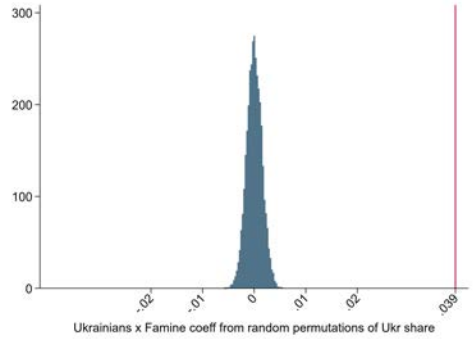
(a) Permute Ukrainian Population Share, Province-Level Panel



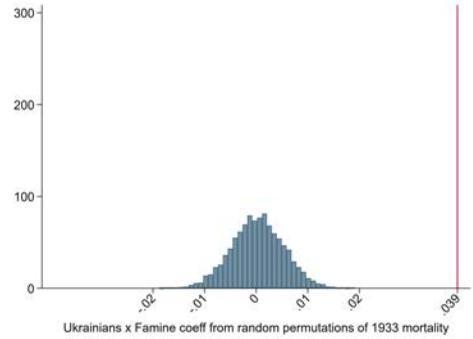
(b) Permute 1933 Mortality, Province-Level Panel



(c) Permute Ukrainian Population Share, District-Level Panel



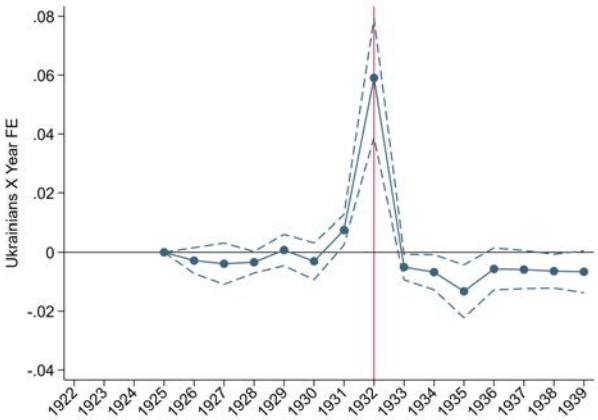
(d) Permute 1933 Mortality, District-Level Panel



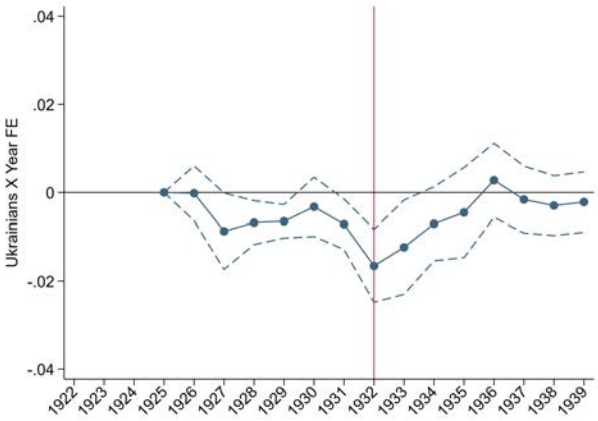
Notes: The figures are histograms of the interaction coefficients ($Ukrainian_i \times Famine_t$). The red vertical lines indicate the baseline estimates reported in the tables.

Figure A.4: Estimates of $Ukrainian \times Year\ FEs$ on Rural Famine Severity

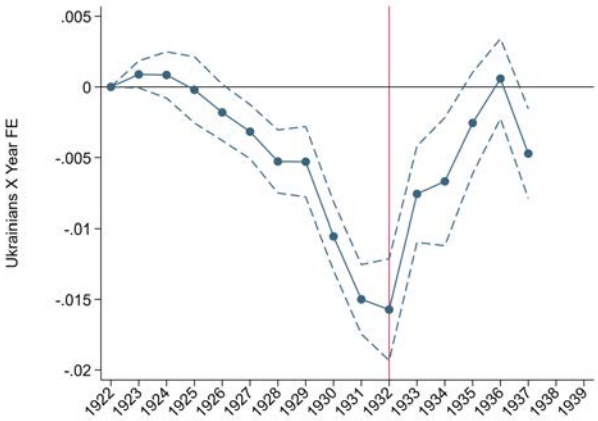
(a) Rural Mortality



(b) Rural Natality



(c) Rural Birth Cohort Size



Notes: Mortality is the number of deaths per 1,000 individuals. Natality is the number of live births per 1,000 individuals. Birth cohort size is the number of individuals born each year divided by the total population as reported by the 1939 census. The dynamic effects figures plot estimates of the interaction coefficients of Ukrainian population share and year dummy variables with their 95% confidence intervals. The estimates and their standard errors are presented in Appendix Table A.4.

Data Appendix

Province-level Panel

The province-level panel includes 1922 to 1940 and 19 provinces within the republics of Belarus, Russia and Ukraine. These provinces correspond to the 1932 administrative division. Belarus and Ukraine were each a single province. The omitted territories are those with no reliable mortality data: Far Eastern Province, Yakut Autonomous SSR, and the North Caucasus ethnic territories: Chechen Autonomous Province, Cherkess Autonomous Province, Dagestan Autonomous SSR., Ingush Autonomous Province, Kabardino-Balkarian Autonomous Province, Karachay Autonomous Province, North Ossetian Autonomous Province. Figure A.1a maps the provinces in our sample. Omitted territories are in white.

Total, Rural and Urban Population For the years 1927, 1937 and 1939, we use data from population censuses.[4] For earlier years, 1920, 1923, 1924 and 1925 we use population estimates constructed by Soviet statisticians (see the exact sources below). For 1933, we use the same source as the 1933 district mortality: in addition to reporting the number of live births and deaths it also reports the population for the beginning of the year that was served by the civil acts registration bureaus (ZAGSy). For the remaining years, we interpolate population. In the paper, we check that our main finding is not an artifact of the interpolation by repeating our estimates with only 1926 population census data.

Administrative boundaries differ across years. However, since the data are reported at disaggregated administrative units, we are able to harmonize the data over time by using ArcGIS and manually aggregate the population data to the 1932 province borders. One issue is that small changes in borders that occur over time lead to large changes in population if we assume that the population is uniformly distributed across space in very large and sparsely populated provinces such as Ural, and West and East Siberia. To address this, we use the 1897 Population Census (the most recent available census prior to the start of our sample), which can be disaggregated to the *Uezd* level (of which there are 817 for the Russian Empire). These data allow us to calculate population density, which we use to attribute population to the 1932 province borders.

The data sources are as follows. 1920: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 1.B. 1922: total population is interpolated between 1920 and 1923; urban population is interpolated between 1920 and 1925. 1923: total population is calculated using the total number of deaths and deaths per 10,000 from Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 5; urban population is interpolated between 1920 and 1925. 1924: total population is calculated using the total number of deaths and deaths per 10,000 from Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 8; urban population is interpolated between 1920 and 1925. 1925: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 1.B. 1926: is interpolated between 1925 and 1927. 1927: De-

[4]The 1926 Census took place on December 17, 1926, all other Soviet censuses took place in January; we use 1926 Census counts for the 1927 population. Population in the 1939 Census is widely believed to be inflated. This was recently corrected by Russian demographers using archival data (Bogoyavlensky, 2013). We use the corrected count.

cember 17, 1926 Population Census. 1928–1932: is interpolated between 1927 and 1933. 1933: Russian state archive of economy (hereafter, RGAE) 1562/329/19 p. 1–12. 1934–1936: is interpolated between 1933 and 1937. 1937: the 1937 Population Census from Zhiromskaya, V.B. and Kiselev, I.N. and Polyakov, Yu.A. (1996) *"Polveka pod grifom "sekretno": Vsesoyuznaya perepis naseleniya 1937 goda [Classified for half a century: All-Union population census of 1937]"*, Moscow: Nauka. 1938: is interpolated between 1937 and 1939. 1939: the 1939 Population Census corrected for the centralized additions (*pripiski*) from Demoscope.ru. 1940: used 1939 value.

Natality and Mortality Mortality is the total number of deaths divided by population (crude death rate). Natality is the number of live births divided by population (crude birth rate). We assign these data into 1932 province boundaries following the same procedure as for population. The rural-urban decomposition of deaths and births is available since 1926.

The data sources are as follows. 1923: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 5. 1924: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part I, Table 8. 1925: Tsentralnoye Statisticheskoye Upravleniye S.S.S.R. [Central Statistical Office of the USSR] (1928) *"Yestestvennoye dvizheniye naseleniya Soyuza S.S.R. 1923–1925 [Natural movement of the population of the USSR]"*, Volume I, Issue 1, Table 1. 1926: Yestestvennoye dvizheniye naseleniya Soyuza S.S.R. v 1926 g, Izdaniye TsSU SS.S.R. (1929), Table 1. 1927–1932: Belarus, Ukraine – RGAE 1562/329/256; Russia – Demoscope.ru. 1933–1940: Demoscope.ru.

Ethnic Composition Ethnic composition comes from the 1897 and the 1926 Population Censuses. The 1897 Census reports population by mother tongue. We use the share of people whose mother tongue is Belorussian, Russian (*Velikorusskiy*), and Ukrainian (*Malorusskiy*). The 1926 Census reports population by self-proclaimed ethnicity and by mother tongue, we use both. Data are calculated in our province borders using 1897 and hand-created district (*volost*)-level 1926 maps. The 1897 map is from Kessler, Gijs and Andrei Markevich, Electronic Repository of Russian Historical Statistics, 18th - 21st centuries, https://ristat.org/, Version I (2020).

Age Structure Region (*okrug*)-level population by 1-year age groups from the 1926 Population Census is reported by Demoscope.ru. We calculated the share of people aged 10 and younger using hand-created region (*okrug*)-level map. This procedure is legitimate because regions (*okruga*) are smaller than our provinces.

Gender Ratio Male to female ratio is from the 1926 Population Census. We calculated it in our province borders using hand-created district (*volost*)-level 1926 map. This procedure is legitimate because districts (*volosty*) are smaller than our provinces.

Grain Harvest, Sown Area, and Yield We map the grain data into 1932 provinces borders following the same procedure as for population. The years 1922, 1924–27 are reported for larger units than our provinces. Thus, we map them into the province borders proportional to the 1913 *Uezd* sown area data.

The data sources are as follows. 1901–1914: Obukhov V.M. (1927) *"Dvizheniye urozhayev zernovykh kultur v Yevropeyskoy Rossii v period 1883–1915 g.g. [Movement of grain crops in European Russia in the period 1883–1915]"* and *Yezhegodnik Rossii* 1904–1916. 1922: Tsentralnoye Statisticheskoye Upravleniye

[Central Statistical Office] (1924) *"Sbornik statisticheskikh svedeniy po Soyuzu S.S.R. 1918–1923. Za pyat let raboty Tsentralnogo Statisticheskogo Upravleniya [A collection of statistical information on the USSR 1918–1923. Five years of work of the Central Statistical Office.]"*, Volume XVIII of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part VI, Tables 7 and 8. 1923: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1924) *"Statisticheskiy yezhegodnik 1922 i 1923 g. (Vypusk pervyy) [Statistical Yearbook 1922 and 1923 (First Issue)]"*, Volume VIII, Issue 5 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part III, Tables 3 and 4. 1924: Tsentralnoye Statisticheskoye Upravleniye [Central Statistical Office] (1926) *"Statisticheskiy yezhegodnik 1924 god (Vypusk pervyy) [Statistical Yearbook 1924 (First Issue)]"*, Volume VIII, Issue 7 of Trudy Tsentralnogo Statisticheskogo Upravleniya [Proceedings of the Central Statistical Office], Part III, Tables 6 and 7. 1925–1927: Statisticheskoye izdatelstvo TsSU SS.S.R. [Statistical Publishing House of the Central Statistical Office of the USSR] (1929) *"Selskoye khozyaystvo SS.S.R. 1925–1928. Sbornik statisticheskikh svedeniy k XVI Vsesoyuznoy partkonferentsii [Agriculture of the USSR 1925–1928. A collection of statistical information for the XVI All-Union Party Congress]"*, Part III. 1928: RGAE 1562/329/1409. 1929–1930: Gosudarstvennoye sotsialno-ekonomicheskoye izdatelstvo [State Socio-Economic Publishing House] (1932) *"Narodnoye khozyaystvo SS.S.R.. Statisticheskiy spravochnik 1932 [The national economy of the USSR. Statistical Handbook 1932]"*, Part II.3.A, Tables 30 and 33. 1931: Gosudarstvennoye izdatelstvo kolkhoznoy i sovkhoznoy literatury "Selkhozgiz" [State publishing house of collective and state farm literature "Selkhozgiz"] (1936) *"Selskoye khozyaystvo SS.S.R.. Yezhegodnik 1935 [Agriculture of the USSR. Yearbook 1935]"*, p. 269, Tables 106 and 107. 1932–1940: RGAE 1562/329/1409.

Procurement We calculated 1925–27 procurement data in administrative borders corresponding to our provinces using hand-created ArcGIS maps (each year is reported using a different administrative division). This operation is legitimate because reported data are more disaggregated than our provinces. 1928–33 data is used as reported.

The data sources are as follows. 1924: Tsentralnoye Konventsionnoye Byuro Khlebozagotoviteley [Central Conventional Bureau of Grain Procurers] (1928) *"Yezhegodnik khlebnoy torgovli N1 [Yearbook of grain trade N 1]"*, Table 6. 1925: Tsentralnoye Konventsionnoye Byuro Khlebozagotoviteley [Central Conventional Bureau of Grain Procurers] (1928) *"Yezhegodnik khlebnoy torgovli N1 [Yearbook of grain trade N 1]"*, Table 14. 1926: Tsentralnoye Konventsionnoye Byuro Khlebozagotoviteley [Central Conventional Bureau of Grain Procurers] (1928) *"Yezhegodnik khlebnoy torgovli N1 [Yearbook of grain trade N 1]"*, Table 22. 1927: Statisticheskoye izdatelstvo TsSU SS.S.R. [Statistical Publishing House of the Central Statistical Office of the USSR] (1929) *"Selskoye khozyaystvo SS.S.R. 1925–1928. Sbornik statisticheskikh svedeniy k XVI Vsesoyuznoy partkonferentsii [Agriculture of the USSR 1925–1928. A collection of statistical information for the XVI All-Union Party Congress]", Part V.* 1928: calculated from the 1928 grain harvest and procurement as a share of harvest from RGAE 4372/30/871 p. 30. 1929: Narodnyy Komissariat Snabzheniya SS.S.R. [People's Commissariat of Supply of the USSR] (1932) *"Yezhegodnik khlebooborota N4 [Yearbook of grain turnover N 4]"*, Tables 3 and 10. 1930: Narodnyy Komissariat Snabzheniya SS.S.R. [People's Commissariat of Supply of the USSR] (1932) *"Yezhegodnik khlebooborota N4 [Yearbook of grain turnover N 4]"*, Table 29 and Table 36. 1931: Komitet po zagotovkam S.-Kh produktov pri SNK SS.S.R. [Committee for Procurement of Agricultural Products under the Council of People's Commissars of the USSR] (1934) *"Yezhegodnik khlebooborota za 1931-32, 1932-33 i predvaritelnyye itogi zagotovok 1933 g. [Yearbook of grain turnover for 1931-32, 1932-33 and preliminary results of procurement in 1933]"*, Table 21. 1932: Komitet po zagotovkam S.-Kh produktov pri SNK SS.S.R. [Committee for Procurement of Agricultural Products under the Council of People's Commissars of the USSR] (1934) *"Yezhegodnik khlebooborota za 1931-32, 1932-33 i predvaritelnyye itogi zagotovok 1933 g. [Yearbook of grain turnover for 1931-32, 1932-33 and preliminary results of procurement in 1933]",* Table 33. 1933: Komitet po zago-

tovkam S.-Kh produktov pri SNK SS.S.R. [Committee for Procurement of Agricultural Products under the Council of People's Commissars of the USSR] (1934) *"Yezhegodnik khlebooborota za 1931-32, 1932-33 i predvaritelnyye itogi zagotovok 1933 g. [Yearbook of grain turnover for 1931-32, 1932-33 and preliminary results of procurement in 1933]",* Table 53.

Grain Targets The First Five-Year plan reports plans for population, grain production, and procurement ("marketable grain"). Aggregate Soviet-level figures are available for every year from 1928 to 1933; from these yearly figures it is clear that the planners just linearly interpolated targets between 1928 and 1933. Disaggregated province-level figures are available for 1928 and 1933; we linearly interpolate planned grain production, procurement and population targets between these years. Summing up the interpolated province targets mechanically results in the aggregate target for that year. Conceptually, the interpolation assumes that the central planners used the same formula to set targets at the aggregate and province levels.

The provinces in the Plan are slightly larger than the provinces in our sample (e.g. West and East Siberia are reported together, Ivanovo and Moscow provinces are united into a Central Industrial Region, Leningrad province and Karelia are united, Tatar and Chuvas regions are united with the Middle Volga region). We calculate grain production and procurement targets in kilograms per person per day (planned grain divided by planned population). Planovoye khozyaystvo [Planned economy] (1930) *"Pyatiletniy plan narodno-khozyaystvennogo stroitel'stva SSSR. Tom 3: rayonnyy razrez plana [Five-year plan for the national economic construction of the USSR. Volume 3: regional aspect of the plan.]"*

Collectivization 1927: Statizdat TSSU SS.S.R. [Statistical publishing house of the Central Statistical Office of the USSR] (1929) *"Kollektivizatsiya Sovetskoy derevni. Predvaritelnyye itogi sploshnykh obsledovaniy 1928 i 1929 gg. [Collectivization of the Soviet countryside. Preliminary results of comprehensive surveys in 1928 and 1929]"*, Table 10. 1928: RGAE 1562/82/271. 1929: Gosplan S.S.S.R. i RSFSR. Ekonomiko-statisticheskiy sektor [State Planning Committee of the USSR and the RSFSR. Economic and statistical sector] (1931) *"Kolkhozy v 1929 g. Itogi sploshnogo obsledovaniya kolkhozov [Collective farms in 1929. Results of a comprehensivy survey of collective farms]"*. 1930: Gosplan S.S.S.R.. Upravleniye Narodnokhozyaystvennogo Ucheta [State Planning Committee of the USSR. Department of National Economic Accounting] (1931) *"Kolkhozy v 1930 g. Itogi raportov kolkhozov k XVI s'yezdu VKP(b) [Collective farms in 1930. Resume of the collective farms' reports to the XVI Congress of the CPSU(b)]"*. 1931: Izd. Kolkhoztsentra SS.S.R. i RSFSR [Publishing House of the Collective Farm Center of the USSR and the RSFSR] (1931) *"Kolkhoznoye stroitelstvo v SS.S.R. [Collective farms building in the USSR]"*, p. 15 and Davies and Wheatcroft (2009), Table 27. 1932: RGAE 1562/82/271. 1933: *"Plan. Zhurnal Gosplana i TsUNKhU SS.S.R. [Plan. Journal of the State Planning Committee and TsUNKhU USSR]"*, 2-1933. 1934–1936: RGAE 1562/82/271. 1937: interpolated between 1936 and 1938. 1938: Gosplanizdat (1939) *"Selskoye khozyaystvo Soyuza S.S.R. 1939 (Statictícheskiy spravochnik) [Agriculture of the USSR 1939 (Statistical handbook)]"*, Part IV.

Dekulakization The baseline measure of kulak households exiled during 1930--31 per 1930 population is estimated as the average between Exiled kulaks (DW) and Exiled kulaks (OGPU) defined below. Exiled kulaks (DW) is the number of *dekulakized* and exiled households in Category II of kulaks in 1930--31 according to Davies and Wheatcroft (2009) (Table 28) per 1930 population. Exiled kulaks (OGPU) is the number of *dekulakized* and exiled households of all categories between January 1, 1930 and July 1, 1931 according to an OGPU (secret police) 1931 report per 1930 population. The report is published in Berelovich A. and V. Danilov (2003). *"Sovetskaya derevnya glazami VChk-OGPU-NKVD. 1918—1939. Documents i materialy"* [Soviet Countryside from the Perspective of VChK-OGPU-NKVD]. Moscow: Rosspen. Vol. 3 "1930—1934 gg.", Book 1. "1930—1931 gg.", document 253. Planned kulaks (lower bound) and

Planned kulaks (upper bound) is the OGPU (secret police) planned number of *dekulakizations* by as of February, 1930 per 1930 population. The planned figures are published in Danilov, Victor, Robert Manning and Lynne Viola (Eds.). (1999-2006). *"Tragediya Sovetskoj Derevni. Kollektivizatsiya i raskulachivanie. Dokumenti i materialy v 5 tomakh, 1927-1939"* [Tragedy of the Soviet Countryside. Collectivization and Dekulakization. Documents and Materials. 5 volumes]. Moscow: Rosspen. Volume 2 "November 1929 — December 1930", Document 69. Total kulaks (OGPU estimate) is the total number of kulaks in the rural population according to the OGPU (secret police) estimate published in Berelovich A. and V. Danilov (2003). *"Sovetskaya derevnya glazami VChk-OGPU-NKVD. 1918—1939. Documents i materialy"* [Soviet Countryside from the Perspective of VChK-OGPU-NKVD]. Moscow: Rosspen. Vol. 3 *"*1930—1934 gg.*"*, Book 1. *"*1930—1931 gg.*"*, document 253. Arrested kulaks 1930 is the number of peasants processed by "troiki" in 1930 per 1930 population according to the OGPU (secret police) estimate published in Berelovich A. and V. Danilov (2003). *"Sovetskaya derevnya glazami VChk-OGPU-NKVD. 1918—1939. Documents i materialy"* [Soviet Countryside from the Perspective of VChK-OGPU-NKVD]. Moscow: Rosspen. Vol. 3 *"*1930—1934 gg.*"*, Book 1. *"*1930—1931 gg.*"*, document 279.

Peasant Resistance to the Soviet Regime "Terrorist acts", unrest demonstrations, and anti-Soviet leaflets registered by the OGPU (secret police) between January 1, 1930 and April 1, 1932 per 1,000 1930 population are according to two OGPU reports. The reports are published in Berelovich A. and V. Danilov (2003). *"Sovetskaya derevnya glazami VChk-OGPU-NKVD. 1918—1939. Documents i materialy"* [Soviet Countryside from the Perspective of VChK-OGPU-NKVD]. Moscow: Rosspen. Vol. 3 *"*1930—1934 gg.*"*, Book 1. *"*1930—1931 gg.*"*, document 272, and Danilov, Victor, Robert Manning and Lynne Viola (Eds.). (1999-2006). *"Tragediya Sovetskoj Derevni. Kollektivizatsiya i raskulachivanie. Dokumenti i materialy v 5 tomakh, 1927-1939"* [Tragedy of the Soviet Countryside. Collectivization and Dekulakization. Documents and Materials. 5 volumes]. Moscow: Rosspen. Volume 3 "Late 1930 — 1933", Document 118.

Peasant Resistance to the Tsarist Regime Peasant revolts in 1895—1914 are from Gokmen and Kofanov (2020).

Bolshevik Votes 1917 Bolshevik vote share is from Protasov et al. (2014). Data is calculated in our province borders using district (*uezd*)-level 1917 map from Castañeda Dower and Markevich (2021).

Communists We calculated 1922 and 1927 data in administrative borders corresponding to our provinces using hand-created ArcGIS maps (each year is reported using a different administrative division). The reported data are more disaggregated than our provinces. For 1931, we use the reported data.

The data sources are as follows. 1922: Izdatelskoye otdeleniye TsK RKP [Publishing Department of the Central Committee of the RCP] (1922) *"Vserosssiyskaya perepis chlenov RKP 1922 goda [All-Russian census of the members of the RCP in 1922]"*, Issue 3, Table 6. 1927: Statisticheskiy otdel TsK VKP(b) [Statistical Department of the Central Committee of the CPSU(b)] (1927) *"Vsesoyuznaya partiynaya perepis 1927 goda. Chislennyy sostav VKP(b) na 10 yanvarya 1927 g. [All-Union Party Census of 1927. The composition of the CPSU(b) on January 10, 1927]"*, Issue 1. 1931: Tsentralnyy Komitet VKP(b). Organizatsionno-instruktorskiy otdel [Central Committee of the CPSU(b). Organizational and instructor department] (1932) *"Sostav VKP(b) v tsifrakh. Dinamika osnovnykh pokazateley rosta parti za 1930 i pervoye polugodiye 1931 g. [Composition of the CPSU(b) in numbers. Dynamics of the main indicators of the growth of the party for 1930 and the first half of 1931]"*

Voting Delegates 1930 We collected location and ethnicity of all 1930 Party Congress delegates that served as province-, district-, city-, or borough-level Party secretary from Rossiyskiy Gosudarstvennyy

Arkhiv Sotsial'no-Politicheskoy Istorii (Russian State Archive of Socio-Political History, RGASPI), Fund 58, Register 1, Files 1–16.

Religious Composition Religious composition is from the 1897 Population Census, available at Kessler, Gijs and Andrei Markevich, Electronic Repository of Russian Historical Statistics, 18th - 21st centuries, https://ristat.org/, Version I (2020).

Shares of Repartition Commune and Private Land Data on commune and private land ownership are originally from the 1905 Land Census. We calculate province shares from district (*uezd*)-level figures taken from Castañeda Dower and Markevich (2018), using manually constructed ArcGIS district (*uezd*)-level maps.

Pre-Soviet Wealth Measures Nominal regional income per capita in 1897, real regional income per capita in 1897, regional labor productivity in 1897, regional rural labor productivity in 1897 (upper and lower estimates) are calculated from corresponding measures for imperial provinces, using hand-created ArcGIS district (*uezd*)-level maps. Imperial province estimates are from Markevich (2019). We estimate the value of agricultural machines by multiplying the number of agricultural machines of different types by their prices and taking the sum. Agricultural machines data are originally from the 1910 Census of Agricultural Machines. We calculate province shares from district (*uezd*)-level figures taken from Castañeda Dower and Markevich (2018), using hand-created ArcGIS district (*uezd*)-level maps. Prices are from Ministerstvo Zemledeliya [Ministry of Agriculture] (1917). *"Sbornik statistiko-ekonomicheskikh svedenij po sel'skomu khozyajstvu Rossii i inostrannikh gosudarstv. [A collection of statistical and economic information about agriculture in Russian and foerign countries]"*, Volume X. Horses, cows, and livestock in 1916 are originally from the 1916 Agricultural Census. We calculate province shares from district (*uezd*)-level figures taken from Castañeda Dower and Markevich (2018), using hand-created ArcGIS district (*uezd*)-level maps.

Grain, Sown area, Yield Obukhov V.M. (1927) *"Dvizheniye urozhayev zernovykh kultur v Yevropeyskoy Rossii v period 1883–1915 g.g. [Movement of grain crops in European Russia in the period 1883–1915]"*.

District-level Panel

District-level dataset includes district for which we were able to collect mortality data for 1933: 1928 and 1933, approximately 1,600 districts within the republics of Russia and Ukraine. Appendix Figure A.1c maps the districts in our sample (omitted territories are in white).

Mortality 1928: Russia: State archive of the Russian federation (GARF) 374/23/7, 13, 31–32, 67, 72–91, 132, 158; Ukraine: Tsentralna Statistichna Uprava USRR [Central Statistical Office of Ukraine] (1929) "Ukraina: Statisticheskiy Schorichnik 1929 [Ukraine: Statistical Yearbook 1929]." 1933: RGAE 1562/329/18–19.

Ethnic Composition Ethnic composition comes from the 1926 Population Census. This census reports population by self-proclaimed ethnicity and by mother tongue, we use both. Data is calculated in our district borders using hand-created district (*volost*)-level 1926 map.

Urbanization 1928: used value from December 1926 Population Census. This census reports district (*volost*)-level rural population and, separately, the population of each urban settlement. To calculate rural and urban population in 1934 administrative borders, we hand-created district (*volost*)-level 1926 map and located all urban settlements on the map. 1933: RGAE 1562/329/18–19.

Gender Ratio Gender ratio is a ratio of males to females according to the 1926 Population Census. To calculate data in 1934 administrative borders, we hand-created district (*volost*)-level 1926 map.

Collectivization Gosplan S.S.S.R.. Upravleniye Narodnokhozyaystvennogo Ucheta [State Planning Committee of the USSR. Department of National Economic Accounting] (1931) "Kolkhozy v 1930 g. Itogi raportov kolkhozov k XVI s'yezdu VKP(b) [Collective farms in 1930. Resume of the collective farms' reports to the XVI Congress of the CPSU(b)]." 1930 districts matched to 1933 districts by name.