Self-Reinforcing Shocks
Evidence from a Resettlement Policy*

Matti Sarvimäki   Aki Kangasharju

February 2010

Abstract

We study the long-term effects of resettling 11 percent of the Finnish population from areas ceded to the Soviet Union during World War II. Our empirical strategy exploits features of the resettlement policy as a source of plausibly exogeneous variation in population growth. The results suggest that a 10 percent increase in population during the war caused an additional 15 percent growth during the next five decades. The growth was driven by migration and it led to the expansion of the non-primary sector in rural locations. The effect is larger for locations connected to the railway network.

JEL Classification: F12, J10, R12, N94

Keywords: Economic geography, agglomeration, migration

*We thank David Card, Steve Gibbons, Ben Faber, Kristiina Huttunen, Guy Michaels, Henry Overman, Panu Poutvaara, Steve Redding, Daniel Sturm, Marko Terviö, Roope Uusitalo, Olof Åslund and seminar participants at HECER, IIES Stockholm University, LSE, Tinbergen Institute and the Nordic Migration Workshop 2008 for helpful discussions and comments. Tiina Hytönen, Nina Intonen and Miikka Rokkanen provided able research assistance. We also thank Ossi Kotavaara, Harri Antikainen and Jarmo Rusanen for kindly constructing the railway indicator. *Sarvimäki* (corresponding author): Government Institute for Economic Research, Finland and SERC, London School of Economics, m.sarvimaki@lse.ac.uk. *Kangasharju*: Government Institute for Economic Research, Finland, aki.kangasharju@vatt.fi.

1 Introduction

A central theme in economics concerns the extent to which one-off shocks can have permanent or even self-reinforcing effects. To illustrate, suppose that a location experiences an exogenous increase in its population. How would such a shock affect the long-term spatial distribution of economic activity? Textbook models provide conflicting answers. A constant returns model with one homogeneous good, fixed capital stock and mobile labor predicts that the shock affects regional structure only temporarily. A constant returns model of small open economies, several goods, immobile factors and imperfect specialization predicts that the shock has a permanent effect. An increasing returns model with monopolistic competition, transportation costs, mobile firms and mobile labor predicts that the shock could be self-reinforcing.[1]

Understanding the empirical relevance of these models has proved a difficult task.[2] The key problem is that population growth is typically endogenous and thus correlations between the growth rates of two periods are unlikely to reveal a causal relationship. Observable exogenous shocks altering location's population are rare. Furthermore, even if suitable variation would be available, some models cannot be falsified. For instance, a well-known property of the new economic geography models following Krugman (1991b) is that a small shock can start a self-reinforcing process. Yet, this holds only under some parameter values, while other parameter values would imply that the regional structure is extremely stable. Importantly, however, evidence of a self-reinforcing shock could not be explained with standard constant returns models.

This paper studies the resettlement of 11 percent of the Finnish population from areas ceded to the Soviet Union during World War II. One element of the resettlement policy was giving land to displaced farmers. At the time, half of the population was employed in agriculture. Thus the policy created an urgent need to acquire vast amounts of land. This task was accomplished by using all suitable publicly owned land and by expropriating the rest from

[1]See, for example, Borjas (2005), Appleyard and Field (2001) and Krugman (1991a).

[2]The relevant empirical literature is too large to be summarized here, see Overman et al. (2003) and Head and Mayer (2004b) for surveys. Recent contributions include, but are not limited to, Ellison and Glaeser (1997), Davis and Weinstein (2002), Head and Mayer (2004a), Hanson and Xiang (2004), Hanson (2005), Amiti and Cameron (2007), Redding et al. (2007), Redding and Sturm (2008), Greenstone et al. (2009) and Ellison et al. (2009).

private farms. The expropriation was based on an explicit and highly progressive schedule. However, an exception was made for the Swedish-speaking parts of Finland, which were exempt from giving up virtually any land. The displaced farmers were not able to choose where the provided land was located.

We exploit these features of the resettlement policy to study the impact of population shocks on later outcomes in rural areas. More precisely, we use the pre-war amount of government owned land, the pre-war size distribution of privately owned farms and the pre-war share of Swedish-speaking population as instruments for wartime population growth.

This empirical strategy hinges on the assumption that the instruments, and the underlying economic forces that gave rise to them, had no direct effect on post-war outcomes. We provide several pieces of evidence supporting this assumption. First, the identifying variation was shaped during the centuries when Finland formed the eastern part of Sweden. We will argue that while the spatial pattern of this variation persisted, revolution in transportation technology and a shift of the economic center from Stockholm to St Petersburg in the early 19th century removed the economic rationale behind these patterns. Second, we show that our approach passes standard falsification exercises. In particular, the instruments do not explain pre-war population growth and the key results remain stable when we use each instrument individually. We also illustrate that the exclusion restriction would have to be violated by an implausibly large magnitude in order to change the conclusions qualitatively. Third, the results survive a battery of robustness checks including controlling for a rich set of pre-war characteristics, excluding outliers and using a variety of sample selection criteria. Thus we argue that the estimates can be interpreted as measuring the causal effects of the resettlement shock.

We find that the shock was self-reinforcing. According to the point estimates, increasing the population by 10 percent between 1939 and 1949 caused roughly 15 percent additional population growth between 1949 and 2000. We show that this growth was driven by migration and that the increase in labor supply was absorbed by the non-primary sector. The results also suggest that the effect was larger among locations connected to the railway network and among locations that had a larger non-primary sector already before the war.

These results contrast with the findings of previous empirical work based on a quasi-experimental setting. In particular, Davis and Weinstein (2002, 2008) show that the massive shock created by the Allied bombings in Japan did not have long-term effects on Japan's city size distribution or the location of specific industries. Brakman et al. (2004) reach similar conclusions in their analysis of the effects of the Allied bombings in Germany. Miguel and Roland (2006) find that the U.S. bombings in Vietnam did not affect the long-run poverty rates, consumption, infrastructure, literacy or population density.

Of course, we are not the first to argue that population shocks have had important long-term consequences. Perhaps the best known example is the historical narrative on the economic and institutional impact of the population decline in the 14th and 15th century Europe (e.g. North and Thomas, 1973). Some econometric studies also reach comparable conclusions. Acemoglu et al. (2009) document an association between the severity of the Holocaust and long-run economic and political outcomes in Russia. Similarly, Bosker et al. (2007, 2008) argue that World War II changed the West German city size distribution.

In comparison to the previous literature, our research design provides several attractive features. First, while the findings based on war-related destruction are highly interesting, it is not clear that bombing campaigns should set-off a self-reinforcing process. After a war, partial survival of the infrastructure (e.g. road networks) and the remaining legal incentives (e.g. property rights) provide strong incentives pushing the regional structure towards the pre-conflict spatial configuration. In contrast, return to the ceded areas was not possible in post-war Finland. Second, we focus on rural areas, whereas most of the previous literature has studied cities. It seems reasonable to assume that agglomeration forces play a smaller in rural than in urban areas. Thus finding strong evidence on their presence also in rural locations provides particularly compelling evidence supporting the relevance of agglomeration effects in real economies. Third, we are able to use three conceptually and geographically distinct sources of exogenous variation. Hence we can check for the robustness of our results in ways that were not feasible in the previous studies. Fourth, we are the first to study the impact of a positive population shock in this context.

Our findings also contribute to the literature studying how open economies absorb labor supply shocks. Following Rybczynski (1955), the question has attracted much attention in the field of international trade. More recently, it has been studied by labor economists as researchers have attempted to understand why even very large immigration flows seem to have little or no effect on native labor market outcomes (e.g. Card, 1990, 2001; Hunt, 1992; Carrington and de Lima, 1996; Friedberg, 2001). Two often proposed explanations are native out-migration and changes in the production structure (e.g. Borjas et al., 1997). Available evidence suggest that immigration either increases native out-migration (Filer, 1992; Frey, 1995; Borjas, 1999) or has a negligible effect (Wright et al., 1997; Card and DiNardo, 2000; Card, 2001). In contrast, we find that exogenous migration flows increased later net migration. Furthermore, current empirical studies tend to find that immigration has only limited impact on production structure (Hanson and Slaughter, 2002; Gandal et al., 2004; Lewis, 2003, 2005), while we find that an exogenous growth of the labor force led to the expansion of the non-primary sector in rural locations.[3]

The remaining of the paper is organized as follows. Section 2 discuses the resettlement in detail. Section 3 presents the data. Section 4 introduces our empirical strategy and Section 5 presents the results. Section 6 concludes.

2 Background

During World War II, Finland fought twice against the Soviet Union. The conflict led Finland to cede over tenth of its territory to the Soviet Union and to evacuate the entire population living in these areas. The evacuation created approximately 430,000 displaced persons corresponding to 11 percent of the total population. The most populous part of ceded areas was the region of Karelia located in the southeastern Finland, while two other ceded areas were located in the extremely sparsely populated Lapland in the north.

The plan of resettling the evacuated population was designed in three pieces of legislation: Rapid Resettlement Act, Land Acquisition Act and Settlement Plan. Those who had derived their principal income from agri-

[3]This finding is relevant also for the literature on structural change. Following Baumol (1967), recent theoretical work has emphasized supply-side reasons for structural change and illustrated how different sectoral productivity growth (Ngai and Pissarides, 2007) or capital deepening (Acemoglu and Guerrieri, 2008) may "release" labor from the old sectors and allow new sectors to expand. While our research design is not suitable for directly testing these closed economy models, we nevertheless find that a supply-side shock hastened the pace of structural change.

culture in the ceded areas were entitled to receive cultivable land in the remaining parts of country. As more than half of the labor force was working in agriculture, this decision had a major effect on the allocation of displaced persons. The displaced farmers were not able to choose their destination. Non-agrarian displaced persons received compensation from their lost property in the form of government bonds and were free to choose their destination areas.[4]

In total, 245,724 hectares of existing cultivated land was used for resettlement and 149,675 hectares was cleared for cultivation (Laitinen, 1995). The land was first taken from the state, municipalities, business corporations, church, other public bodies, land speculators and landowners not practicing farming. However, "secondary sources" – private landowners who lived in their farms – ended up providing roughly half of the cultivated fields. The land was purchased either on voluntary basis or through expropriation using a progressive scale presented in Figure 1. Landowners were paid a "justifiable current local price" for the expropriated land in the form of government bonds. However, like all capital owners, they were subject to a large capital tax (which they could pay using these government bonds) and thus did not receive much compensation in practice. That is, the expropriation did not inject cash to the affected municipalities.

The amount of land available for displaced farmers within the borders of a given municipality – and hence the amount of displaced farmers allocated to the municipality – was primarily determined by the pre-war farm size distribution and the amount of land owned by the public sector. Two other factors created variation in the inflow of displaced persons. First, no-one was settled to northern Finland, where the conditions for agriculture are the least favorable. Second, Finland is a bilingual country and the Land Acquisition Act included a clause demanding that the resettlement should not alter the balance of languages within municipalities. Since the vast majority of the displaced farmers spoke Finnish as their mother tongue, very few received land from the Swedish-speaking parts of the country.

As we discuss in detail in Section 4, we will use these features of the resettlement policy to evaluate the causal impact of one-off population shocks.

[4]The only exception was the capital, Helsinki, where the housing shortages led to direct regulation. In 1945, those who wished to move to Helsinki had to apply for a specific permission from the local housing board. The allocation of land to displaced farmers was completed by the end of year 1948.

The plausibility of this empirical approach depends on the reasons why some locations were endowed with more large farms, more government owned land or a larger Swedish-speaking population. The origins of this variation go back to the time when Finland formed the eastern part of Sweden.[5] At the time, most of the economic activity took place in southwestern part of the country, which was well connected to Stockholm by the Baltic Sea. A large fraction of the farmland, and virtually all manors, were located in this area. Over time, permanent settlement expanded towards east and north. A considerable number of migrants from Sweden also settled in the western and southern coasts. Yet, the vast areas farther east and north remained distant hinterland, where people lived off burn-beat cultivation and hunting. These areas became state property in the 16th century as the crown laid claim to the wilderness and actively encouraged colonization in an attempt to increase tax revenues.

In short, the pattern of large farms in the southwest, government owned land in the north and east, and Swedish-speaking settlements in the coasts were present already in the Middle Ages. While this division faded over time, there were still clear differences in the 1930s. Figure 2 illustrates these patters. The bottom-right panel of the figure also presents the population share of the displaced in 1948. While the proportion of displaced persons in many Swedish-speaking municipalities on the western coast is markedly low, municipalities elsewhere experienced up to a third increase in their populations. However, there is also large variation between neighboring municipalities in the Finnish-speaking area.

While these patterns persisted over centuries, the economic forces giving rise to them virtually disappeared over time. One reason was the rapid population growth and the end of the Little Ice Age, which pushed permanent settlement towards east and north.[6] The second important change was the shift of the political and economic center from Stockholm to St Petersburg in 1809 when Sweden lost Finland to Russia. Even within Finland, the capital

[5]This discussion draws from Kirby (2006). Swedish rule started at around mid-12th century and ended in 1809. Throughout we use "Finland" to refer to the area falling within the 1939 borders.

[6]Between mid-18th and mid-19th century, Finland experienced a roughly 1,5% annual population growth. Little Ice Age refers to the period of global cooling between 16th and mid-19th century. While researchers do not agree on the exact timing of this period, there is a wide consensus on that conditions for agriculture in northern and eastern parts of Finland improved substantially from mid-18th century onwards.

city was moved eastwards from Turku to Helsinki. The third factor was the improvement in transportation technology, in particular the construction of an ambitious railway network starting in 1862. Even though the market area of St Petersburg disappeared with the Russian revolution and the consequent Finnish independence in 1917, the economic center did not return to the southwest.

3 Data

Our primary source of information are various Statistical Yearbooks and Agricultural Censuses published by Statistics Finland since the 1930s. These sources provide information at the level of the local administrative unit (municipality). Statistics Finland has provided additional data on municipalities' production structure in 1980, 1990 and 2000 and annual migration flows, fertility and mortality for the period between 1951 and 1997. The data is further augmented with an indicator variable for the municipality being connected to the railway network in 1939 as documented by historical engine driver timetables (see Kotavaara et al., 2010, for details). In order to ensure that the spatial units remain stable over time, we have aggregated all municipalities that either merged or dissolved between 1930 and 2000. The procedure and the data sources are discussed in detail in the Appendix.

In our baseline analysis, we focus on those 349 rural municipalities that did not cede territory to the Soviet Union.[7] Partly ceded municipalities are excluded, since we cannot construct consistent time series for them. The motivation for excluding cities is that our identification strategy relies on instruments that are relevant only for rural areas. In the Appendix we show that the results are not sensitive to this sample selection rule.

Figure 3 plots the population growth rates between 1949 and 2000 on the growth rates between 1939 and 1949. The figure reveals that some municipalities experienced very large changes in their populations and that there is a strong positive association between wartime growth rates and later growth rates. Furthermore, while almost all rural municipalities grew during the

[7]Our definition of rural area is based on Statistics Finland's pre-war categorization, see the Appendix for details. Municipalities in the baseline sample had a median land area of 417 square kilometers and the median population of 4,273 in year 2000. In comparison, counties in the United States had a median land area of roughly 1,600 square kilometers and a median population of 25,000 in year 2000.

war and its immediate aftermath, three quarters lost population during the next five decades. This decline was driven by emigration and, more importantly, urbanization. While urbanization had begun already before the war, there were particularly large migration flows away from rural areas during the late 1960s, early 1970s and late 1990s. In total, the share of the Finnish population living in the baseline sample area decreased from more than two thirds in 1930 to roughly half in 2000.[8]

4 Empirical Strategy

Above we saw that wartime population growth was positively correlated with later population growth. However, this association could be a result of confounding factors affecting both the wartime and the post-war population growth. We next set up a simple empirical framework, which allows for assessing the causal effect. We will estimate the equation

$$g_{ja} = \alpha g_{jw} + X_j \beta + \epsilon_j \tag{1}$$

where g_{ja} is the population growth rate during a years after the end of the resettlement in 1949, g_{jw} is the population growth rate between 1939 and 1949, X_j is a vector of observable characteristics measured before the war and ϵ_j summarizes unobserved factors affecting population growth. The parameter of interest is α. Note that if $\alpha = -1$, shocks undo themselves in a years. Alternatively, if $\alpha = 0$, when a is large, shocks are permanent. If $\alpha > 0$, shocks are self-reinforcing.

The challenge in consistently estimating α is that population growth in both periods may be affected by unobserved factors. The direction of this bias is not known *a priori*. Typically one would expect people to migrate into locations with higher growth potential and thus to bias OLS estimates upwards. However, it is not evident that this was the case in Finland during the 1939–1949 period. At the time many influential policy makers argued that national security required self-sufficiency in food production and a more evenly distributed population (Laitinen, 1995; Pihkala, 1952). These concerns led to policies encouraging people to settle in areas, which turned out

[8]This calculation for 1930 excludes areas that were later ceded to the Soviet Union also for the pre-war period.

to have low growth potential.[9] Hence ordinary least-squared estimates of (1) could also be biased downwards.

We address the issue in two ways. First, we control for pre-war observable characteristics (growth trend, mean income, production structure, population density, a market potential measure, an indicator for being connected to the railway network) and constant geographical characteristics (longitude, latitude, proximity to a city). While these variables may not capture all factors affecting wartime population growth, we expect conditioning on them to reduce the potential bias.

Our main identification strategy, however, is to use an instrumental variables approach exploiting the three elements of the allocation policy discussed above. The instruments are the share of municipality's population speaking Swedish as their mother tongue in 1930, hectares of publicly owned land per capita in 1940 and hectares of privately owned expropriable agricultural land per capita.[10] Thus the key identifying assumptions are (a) that these instruments were uncorrelated with post-WWII spurious shocks and (b) that they had no direct effect on population growth. Since the instruments are outcomes of long historical processes, they cannot be affected by post-war shocks. Above we have also argued that the reasons giving rise to the variation in the instruments had lost their relevance by mid-20th century. We also note that while Finland has practiced varying regional policies in the post-war period, these policies have not depended on the number of displaced farmers each region received. In the next section, we will use the available data to provide further evidence supporting this identifying assumption.

[9]In particular, the displaced persons were offered the option receiving a "cold farm" instead of already cultivated agricultural land. "Cold farms" were located in eastern and northern Finland and had no cultivated land or buildings. Thus the receiver had to clear the land and found the farm by himself, while the state provided special awards and payment arrangements. However, take-up for this option was low. (Waris et al., 1952)

[10]We approximate the available privately owned agricultural land by using the expropriation scale presented in Pihkala (1952, Table II; reproduced in Figure 2) and the 1930 size distribution of privately owned land. Specifically, the instrument is constructed as $I_{i39} = \sum_{s=1}^{n} (\tau_l^s h_l^s + \tau_m^s h_m^s)\, N_{i30}^s / P_{i39}$, where τ_l^s is the expropriation rate at the lower limit of the size class s, τ_m^s is the expropriation rate for the part exceeding the lower limit in this bracket, h_l^s is the bracket's lower limit in hectares, h_m^s is the midpoint of the exceeding part, N_{i39}^s is the pre-war number of farms in the municipality belonging to the bracket in municipality i, and P_{i39} is municipality's 1939 population.

5 Results

We start with a falsification exercise. The first two columns of Table 1 report the results of regressing the population growth between 1930 and 1939 on the instruments. This exercise is motivated by the assumption that if the instruments had a direct impact – or were correlated with unobserved factors that had on an impact – on post-war population growth, they would also be associated with pre-war growth. However, we do not find such correlation from the data. The only statistically significant estimate is the negative association between the availability of privately owned land and pre-war population growth, suggesting that the second-stage estimates would be biased downwards (see the Appendix for discussion). This correlation disappears once we condition on basic observable municipality characteristics measured in 1930.

In contrast, the first-stage estimates reported in the third and fourth columns of Table 1 reveal a strong association between wartime population growth and the instruments. Consistent with the settlement plan, larger stock of available agricultural land is positively correlated with population growth between 1939 and 1949. Similarly, municipalities with large Swedish-speaking population share received fewer displaced persons and thus grew less. Together the instruments explain roughly a sixth of the variance in the wartime population growth. The estimates are very similar in the baseline specification and in the specification controlling for population growth between 1930 and 1939, share of labor force in primary sector in 1930, mean taxable income per capita in 1939, population density in 1939, a market access measure for 1939, indicators for being a neighbor of a city and being connected to the railway network in 1939, and longitude and latitude.[11] The F-statistics imply that the estimates should not suffer from problems related to weak instruments.

The finding that the instruments do not explain pre-war growth, but have a strong impact on wartime growth is in line with our identifying assumption. In the Appendix, we report further results supporting our approach. In particular, we find that the instruments yield similar results when used

[11]We measure market access as a log distance weighted sum of the 1939 population of all municipalities. We follow the original formulation of Harris (1954) and use (Euclidean kilometer distance)$^{-1}$ as weights. The market access measure for 1930 is constructed similarly as a distance weighted sum of the 1930 population.

individually. Furthermore, we show that the exclusion restrictions would have to be violated by an implausibly large magnitude in order change the results qualitatively. We also discuss results using alternative sample areas and subsamples where we gradually exclude the most influential observations (outliers). The results are remarkably stable across specifications. Thus we interpret the estimates reported next as measuring a causal relationship.

Table 2 reports the main results. The estimates come from 24 separate regressions that differ in the length of the post-war period studied, the estimation method used and the inclusion of control variables. The first column of panel A reports OLS (first row) and 2SLS (second row) estimates from regressing population growth between 1949 and 1950 on the population growth between 1939 and 1949. Similarly, the sixth column reports the estimates from regressing population growth between 1949 and 2000 on population growth between 1939 and 1949. Panel B reports corresponding estimates after controlling for pre-war municipality characteristics and geographical indicators.

Recall that if the resettlement shock had a temporary effect, the estimates should approach minus unity as we extend the study period. Clearly, this does not occur. Rather, the estimates are positive and become larger when the study period is extended. The point estimates reported in the sixth column of panel B imply that an exogenous migration flow increasing municipality's population by 10 percent during the war led to a further population growth of roughly 15 percent during the next five decades. All estimates are statistically highly significant.[12] Thus the results strongly suggest that wartime population shocks were self-reinforcing.

Having established a link between wartime and post-war population growth, we next ask what drove this growth. We note that migration of workers and firms is the key mechanism of the new economic geography models motivating our study. However, differences in population growth could also follow from differences in fertility or mortality. To assess the importance of these channels, we regress measures of post-war migration flows, fertility and mortality on wartime population growth. Table 3 reports the results.

[12] We report conventional standard errors because they are larger than the robust standard errors in part of the specifications. We take this as an indication of the robust standard errors potentially suffering from small sample bias. In the Appendix, we show that the choice of inference approach, including an approach allowing for spatially correlated error terms, does not affect our conclusions.

In the first column, the dependent variable is the sum of municipality's annual net migration flows between 1951 and 1997 scaled by its population in 1951. The estimates indicate that the resettlement shock increased later net migration. Interestingly, when we study in- and out-migration separately, we find positive effect on both. The impact on out-migration is consistent with an earlier result showing that the resettlement increased geographical mobility among the displaced also in the post-war period (Sarvimäki et al., 2009). It is also consistent with the hypothesis that many "natives" could have responded to the influx of displaced people by moving out. Yet, the impact on inflows is even larger.

We also find a positive effect on fertility. However, these differences do not follow from the displaced people having higher fertility rates than the rest of the population. Rather, the positive impact on fertility is likely to reflect the fact that migrants are typically in their prime child-bearing age.[13] Thus we interpret the impact of wartime population shocks on fertility to be a consequence of increased migration.

Why did so many people decided to migrate into these areas? The most likely explanation is that wartime population growth improved economic opportunities. Economic models leading to this prediction typically assume non-negligible transportation costs for intermediate and final products, benefits arising from the size of the labor market (due to better matches or insurance against idiosyncratic shocks) or knowledge spillovers. While the available data do not allow us to assess the relative importance of these agglomeration forces, it seems reasonable to think that they are more relevant for the non-primary than for the primary sector.

Results presented in Table 4 support this reasoning. Panel A reports the estimates from regressing the growth of labor force working in primary sector on wartime population growth and the same pre-war control variables as above. Panel B reports similar estimates using growth of labor force working in non-primary sector as the dependent variable. The results reveal

[13]We reach this conclusion using individual-level data drawn from the 1950 and 1970 censuses, see Sarvimäki et al. (2009) for details. Regressing the number of children in 1950 on a dummy for the person living in the ceded area in 1939 yields an estimate of −0.04 (standard error 0.03). A similar regression with the number of children in 1970 also yields a point estimate −0.04 (standard error 0.01). In contrast, regressing the number of children in 1970 on a dummy for the person having changed the municipality of residence between 1950 and 1970 yields an estimate of 0.48 (standard error 0.01). Once we control for age, the estimate becomes −0.04 (standard error 0.01).

that the resettlement shock had virtually no impact on the size of the primary sector – at least after year 1950, when we first observe post-war production structure. However, we find a strong positive impact on the growth of non-primary sector. Hence the resettlement shock appears to have been absorbed by the non-primary sector.

Our last question concerns the role of initial conditions. Table 5 reports the results when we interact the wartime population growth with an indicator variable taking value one if municipality's share of labor force working in the primary sector in 1930 was above 88 percent (sample median) and zero otherwise. Table 6 reports similar estimates when initial conditions are measured as a connection to the railway network in 1939. Both specifications suggest that initial conditions mattered. According to the point-estimates, having above median non-primary sector or an access to efficient transportation system amplified the impact of population shocks. However, the 2SLS estimates are imprecise and most of the estimates for interactions are not statistically significant. Hence we take this evidence to be more suggestive than the results discussed above. Nevertheless, it is tempting to hypothesize that municipalities with more favorable initial conditions were more able to expand their modern sectors and thus to benefit more from the increase in their labor force.

6 Conclusions

In this paper, we have studied the long-term impact of resettling more than a tenth of the Finnish population from areas ceded to the Soviet Union during World War II. This quasi-experimental setting allows us to construct plausible instruments predicting wartime population growth among rural municipalities. The instruments exploit the design of the resettlement policy, which allocated displaced farmers to Finnish-speaking municipalities endowed with publicly owned land or large private farms suitable for expropriation. We report several pieces of evidence supporting the validity of this empirical strategy.

The results suggest that a shock increasing municipality's population by 10 percent caused an additional 15 percent population growth during the next five decades. The results are statistically highly significant and survive a battery of robustness checks. Furthermore, we find that the post-war population growth was driven by migration and that the increase in labor supply was absorbed by non-primary sector. Locations with favorable initial conditions – measured as a larger non-primary sector or being connected to the railway network before the war – appear to have been more responsive to the increase in their labor force.

A Robustness Checks

A.1 Validity of the Instruments

In this section, we provide further evidence supporting our key identifying assumption. We also discuss the sensitivity of our broad conclusions to the violations of this assumption. To clarify the issues involved, it is convenient to write the model in matrix notation as

$$\begin{aligned} G_{t+1} &= \alpha G_t + X\beta + Z\gamma + u \\ G_t &= X\beta + Z\pi + v \end{aligned} \tag{A1}$$

where G_{t+1} is a vector of the post-war growth rates, G_t is a vector of the wartime growth rates, X is a matrix of control variables and Z is a matrix of instruments. The parameter of interest is α. In the simple case where we omit X, the probability limit for the 2SLS estimate is

$$\text{plim } \hat{\alpha} = \alpha + \text{plim} \left(G_t' P_Z G_t \right)^{-1} G_t' Z\gamma \tag{A2}$$

where $P_Z = Z \left(Z'Z \right)^{-1} Z'$. Our key identifying assumption is that all elements of parameter vector γ are zero.

Note that in the case of one instrument, the formula for asymptotic bias simplifies to γ/π. From the first-stage regressions (Tables 1 and A1), we know that $\pi_1 > 0$, $\pi_2 > 0$ and $\pi_3 < 0$, where π_1, π_2 and π_3 are the first-stage coefficients for exproriable privately owned land, government owned land and share of Swedish-speakers in the municipality, respectively. Thus, our conclusions could be misguided if the presence of large private farms or large amounts of government owned land – or some unobservable factors correlated with them – had a sufficiently large positive direct effect on post-war population growth. Similarly, a sufficiently large negative direct effect of a Swedish-speaking population could lead to a qualitatively wrong conclusion.

The data provide no support for the availability of land having a positive impact on population growth. If anything, the estimates for pre-war population growth presented in Tables 1 and A1 suggest that the land instruments would be negatively associated with population growth and thus bias the second-stage estimates downwards. On the other hand the point estimate for Swedish-speaking population is negative though the point estimate is small in magnitude and the association is not statistically significant.

We next study whether different instruments lead to different conclusions. Table A2 reports results when we use one instrument at a time. For reference, also the 2SLS estimates using all instruments are reported. On balance, the instruments yield similar estimates across the specifications. The only notable exception are the estimates up until 1970 in the specification where we use expropriable privately owned land as an instrument and omit the control variables. As the first column of Table A1 reveals, this specification is also the only one where we find statistically significant association with pre-war population growth. Since this association is negative, we should expect the second-stage estimates to be biased downwards. Consistent with this reasoning, 2SLS estimates from this specification are smaller than from other specifications. However, when we control for pre-war characteristics, the first-stage association between pre-war population growth and the exproriable privately owned land disappears and the second-stage estimates become similar to those obtained with the other instruments.

The stability of the estimates is quite remarkable given that one might expect the impact of population shocks to vary between municipalities. Indeed, this is what the results presented in Table 5 suggest. Note that if treatment effects are heterogeneous, different instruments identify different weighted averages of local average treatment effects (Angrist and Imbens, 1995). As illustrated by Figure 2, each instrument affects a very different part of the country.[14] Yet, when we control for pre-war characteristics and geographical indicators, the Sargan test rejects only when the dependent variable is population growth between 1949 and 1960.[15]

The finding that all instruments individually lead to the same conclusion is reassuring. However, it is impossible to definitely rule out that all of the instruments would yield similarly biased estimates. To be clear, we find this very unlikely. Nevertheless, it is informative to ask how large the violations of the exclusion restriction would need to be in order to qualitatively change the conclusions.

[14]An alternative way to see that the instruments generate independent variation is to note that they are only weakly correlated with each other: the correlation coefficient between the two land instruments is 0.22, while the correlation coefficient between the share of Swedish-speaking population and privately owned land (publicly owned land) is 0.01 (-0.06).

[15]Note that in heterogeneous response framework, rejection of over-identification tests can be interpreted as evidence of effect heterogeneity. Alternatively, the rejection could follow from one or all instruments being invalid.

Note that if we knew that the true value of γ is γ_0, we could consistently estimate α from

$$(G_{t+1} - Z\gamma_0) = \alpha G_t + X\beta + u \tag{A3}$$

with 2SLS using Z as instruments for G_t (Conley et al., 2008). Of course, we do not know γ_0. Yet, we can perform sensitivity analysis by studying the implications of different assumptions about its values.

Table A3 reports the results for assuming a range of values for γ_0. The dependent variable in all regressions is population growth between 1949 and 2000 and we control for pre-war municipality characteristics. The results illustrate that the conclusions are qualitatively unchanged even in the presence of substantial violations of the exclusion restriction. In order to get a point estimate of zero, one would need to have a prior such as $\gamma_1 = .3$, $\gamma_2 = .3$, $\gamma_3 = -.3$. In words, one would need to assume that an additional hectare of exproriable land per capita directly increased post-war population growth by 30 percent, that the impact of an additional hectare of publicly owned land per capita would have had a similar effect *and* that having an entirely Swedish-speaking population would have decreased population growth by 30 percent even in the absence of the settlement policy.

Given that we find no positive association between the pre-war population growth and the land instruments, it seems very unlikely that these instruments had a strong direct positive impact on post-war population growth. On the other hand, the point estimates presented in Table 1 and A1 suggest a slight negative association between pre-war population growth and municipality's share of Swedish-speaking population. Of course, these estimates are not statistically significant and their sign may be determined purely by chance. Nevertheless, we note that in order to drive the point estimate of α to zero without assuming a positive direct effect of the land instruments, one would need to assume that having a completely Swedish-speaking population decreased the post-war population growth by 60 percent. This seems extremely unlikely. Furthermore, our conclusions remain unchanged when we exclude the Swedish-speaking areas from the estimation sample and use only the land instruments.

A.2 Outliers

Another potential concern is that our results could be driven by some municipalities experiencing extreme population growth or decline. To study this possibility, we gradually exclude observations that are particularly influential for the estimates. We do this by calculating Cook's (1977) distance measure

$$D_i = \frac{\sum_{j=1}^{n} \left(\hat{g}_{j,49+a} - \hat{g}_{j(-i),49+a} \right)^2}{(k+1)\, s^2} \tag{A4}$$

where $\hat{g}_{j,49+a}$ is the prediction from the full regression model for observation j, $\hat{g}_{j(-i),49+a}$ is the prediction for observation j from a refitted regression model in which observation i has been omitted, s^2 is the estimated root mean square error and k is the number of parameters in the model. The intuition of this measure is that it is informative about the influence of data point i for the least squares regression estimate. A common rule of thumb is that an observation is suspected as being an outlier when $D_i > 4/(n-k-1)$. In our case, the threshold value is $D_i > 4/(349-10-1) = .0118$.

Table A4 reports the results when we gradually exclude the most influential observations. For reference, panel A present the baseline estimates in a specification with control variables. Panel B presents the estimates when we exclude observations that have Cook's D values larger than twice of the rule of thumb threshold $.0118 * 2 = .0236$. Panel C reports the estimates from a sample excluding the observations above this threshold value, while panel D and E present the results from excluding observations with Cook's D values larger than a half and a quarter of the rule of thumb value for outliers. All estimates suggest that wartime population growth had a statistically significant positive effect on post-war population growth.

A.3 Alternative Study Areas

As another robustness check, we study whether the results are sensitive to the choice of the study area. Figure A5 shows three reasonable choices for the estimation sample. Our baseline sample, illustrated by the gray area, consists of all rural municipalities that did not loose territory to the Soviet Union. That is, only areas including cities as well as the ceded and partly ceded areas in the eastern part of the country are excluded. A potential concern on using this sample is that it includes northern municipalities which were endowed with plenty of government owned land, but were not influenced by the settlement policy. Hence we repeat the analysis using a restricted sample that excludes the northern part of the country. Next, we repeat the analysis using only those rural municipalities that were mentioned in the Settlement Plan, illustrated by the dotted line in Figure A1.

Table A5 reports the results. For reference, panel A reports the baseline estimates. The conclusions remain unchanged when using the restricted sample (panel B), though the standard errors increase and the point estimates decrease slightly. Regressions using only municipalities mentioned in the Resettlement Plan yield similar results (panel C). Finally, panel D reports the estimates using the baseline data augmented with cities. Again, the results are very similar, though less precise, as those obtained from the baseline sample.

A.4 Alternative Standard Errors

Our final robustness check concerns inference. In Tables 1 to A5, we have chosen to report standard errors based on the assumption of the error terms being homoscedastic and spatially independent. This choice was driven by two factors. First, heteroscedastic robust standard errors are biased downwards when heteroscedasticity is relatively modest and the sample size is small (Chesher and Jewitt, 1987). As it turns out, heteroscedastic robust standard errors are smaller than the conventional standard errors for our key 2SLS estimates. Thus robust standard errors are likely to overstate the precision of these estimates. On the other hand, we show below that allowing for spatial dependence has little effect on the instrumental variables standard errors. However, the issue of computing them requires discussion, which is more convenient to conduct in an Appendix.

Table A6 presents the key estimates and standard errors from alternative approaches. For reference, the first two rows of Panel A and B reproduce the OLS and 2SLS estimates and conventional standard errors reported in Panel B of Table 2. The third rows report heteroscedasticity robust standard errors.[16] For OLS estimates, robust standard errors are substantially

[16]By conventional standard errors we mean square roots of the diagonal elements of the covariance matrix estimator $\hat{V}_c = (X'X)^{-1}\left(\sum \hat{\epsilon}_i^2/N\right)$, where $\hat{\epsilon}_i = y_i - X_i'\hat{\beta}$ is the estimated regression residual. Similarly, heteroscedastic robust standard errors refer to square roots of the diagonal elements of the covariance matrix estimator $\hat{V}_r = N\,(X'X)^{-1}\left(\sum X_i X_i' \hat{\epsilon}_i^2/N\right)(X'X)^{-1}$.

larger than the conventional ones. However, in the 2SLS regressions most of the robust standard errors are smaller than the conventional ones. Given the premium we put to the 2SLS estimates due to their more plausible identification, it seems reasonable to think that conventional standard errors provide more conservative inference than robust standard errors.

We next consider the implications of relaxing the assumption of spatial independence. This assumption would be violated, for example, if a new manufacturing plant launched in one municipality would affect demand for labor and intermediate inputs also in the neighboring municipalities. To allow for such spatial dependence, we use an approach suggested by Conley (1999). This approach is similar to the time series heteroskedasticity and autocovariance (HAC) consistent covariance matrix estimation. More precisely, the spatial GMM estimator takes the same form as GMM estimators for time series or independent data, except that inference and the weighting matrix are based on the covariance matrix estimator

$$\hat{V}_s = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} K_N \left(s_i, s_j\right) \cdot z_{si} \hat{\epsilon}_i z_{sj}^{'} \hat{\epsilon}_j \tag{A5}$$

where $K_N \left(s_i, s_j\right)$ is an uniform kernel taking value one if locations s_i and s_j are within a cut-off distance of each other and zero otherwise. That is, the unobserved factors affecting the outcomes of each municipality are allowed to be correlated among all municipalities within a prespecified distance. While geographic distance is unlikely to be a perfect measure of the true "economic distance" determining spatial dependence, the covariance matrix estimator (A5) remains consistent given that measurement error in the distance proxy is bounded (see Conley, 1999, 2008, for details).

Panel C presents spatial GMM results for the cut-off distances of 50, 100 and 250 kilometers. Both the point estimates and standard errors are similar to those obtained from 2SLS. We also present similar standard errors for OLS estimates in the three last rows of Panel A. These standard errors are close to the heteroscedasticity robust OLS standard errors. Most importantly, all approaches yield statistically highly significant estimates and lead to the same conclusion.

B Data Appendix

B.1 Data sources

Population 1930: Statistics Finland (1979): *Väestön elinkeino 1880-1975.* **Population 1939–2000**: Statistical Yearbooks, various years. **Industry structure 1930–1970**: Statistics Finland (1979): *Väestön elinkeino 1880-1975.* **Industry structure 1980–2000**: Statistics Finland's aggregation from microdata (*Työssäkäyntitilasto*) recording the sector of employment for the entire population living in Finland at the end of each year. **Migration**: Statistics Finland (*pc-Axis, Population Structure*). **Taxable Income per capita** (defined as the taxbase of the municipality, number of *veroäyri*, divided by the population): Central Statistical Office: SVT XXXI A:15. **Swedish speaking population in 1930**: Statistical Yearbook 1937**.** **Longitude and latitude**: Polygon centroid of the municipalities presented in Figure 2. **City**: Statistics Finland / Central Statistical Office categorizes municipalities into cities, market towns and rural municipalities. Our definition of urban area is based on the pre-war category of cities augmented with two municipalities (Espoo and Vantaa) bordering Helsinki (the capital). The municipalities classified as urban are: Helsinki, Espoo, Vantaa (formerly Helsingin maalaiskunta), Tampere, Turku, Vaasa, Lahti, Oulu, Kuopio, Kotka, Kemi, Pori, Lappeenranta, Mikkeli, Rauma, Hämeenlinna, Jyväskylä, Kokkola, Savonlinna, Hanko, Porvoo, Kajaani, Pietarsaari, Joensuu, Hamina, Loviisa, Tammisaari, Iisalmi, Raahe, Uusikaupunki, Heinola, Kristiinankaupunki, Tornio, Kaskinen, Uusikaarlepyy and Naantali. **Neighbouring a city** (municipalities bordering the cities defined above): manual inspection of a map. **Municipality has a railway station**: Kotavaara et al. (2010). **Municipality included in the Settlement Plan**: Paukkunen, L. (1989). *Siirtokarjalaiset nyky-Suomessa.* University of Jyväskylä. **Number of displaced persons in municipalities**: National Archives, SM / Siirtoväenasiainosasto / Kansiot H1, H1a, H2, H3, H4, H5, H6, H7, H8, H9.

B.2 Aggregation

The spatial unit of all variables is a municipality. In order to ensure that the spatial units remain stable over time, we have aggregated all municipalities that either merge or split over the study period. That is, if municipality A merged to municipality B during the study period, we aggregate A and B also for the pre-merging period. Similarly, if municipality C dissolved to D and E, we aggregate D and E also for the post-dissolution period. There are a few instances where a municipality has merged into several municipalities. In these cases, we divide the merging municipality to the host municipalities using host municipalities population (measured one year prior to the merge) as weights. That is, if part of municipality F was merged to G and the rest to H, we assign the share $P_G/\left(P_G+P_H\right)$ to municipality G and $P_H/\left(P_G+P_H\right)$ to municipality H.

References

Acemoglu, D. and V. Guerrieri (2008). Capital Deepening and Nonbalanced Economic Growth. *Journal of Political Economy 116*(3), 467–498.

Acemoglu, D., T. Hassan, and J. Robinson (2009). Social Structure and Development: A Legacy of the Holocaust in Russia. Mimeo.

Amiti, M. and L. Cameron (2007). Economic Geography and Wages. *The Review of Economics and Statistics 89*(1), 15–29.

Angrist, J. and G. Imbens (1995). Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association 90*(430), 431–442.

Appleyard, D. and A. Field (2001). *International Economics* (4th ed.). McGraw-Hill.

Baumol, W. (1967). Macroeconomics of Unbalanced growth: the Anatomy of Urban Crisis. *American Economic Review*, 415–426.

Borjas, G. (1999). *Heaven's door: Immigration policy and the American economy*. Princeton University Press.

Borjas, G. (2005). *Labor Economics* (3rd ed.). McGraw-Hill.

Borjas, G., R. Freeman, and L. Katz (1997). How Much Do Immigration and Trade Affect Labor Market Outcomes? *Brookings Papers on Economic Activity*, 1–90.

Bosker, M., S. Brakman, H. Garretsen, and M. Schramm (2007). Looking for Multiple Equilibria when Geography Matters: German City Growth and the WWII Shock. *Journal of Urban Economics 61*(1), 152–169.

Bosker, M., S. Brakman, H. Garretsen, and M. Schramm (2008). A Century of Shocks: The Evolution of the German City Size Distribution 1925–1999. *Regional Science and Urban Economics 38*(4), 330–347.

Brakman, S., H. Garretsen, and M. Schramm (2004). Bombing of German Cities during WWII and its Impact on City Growth. *Journal of Economic Geography 4*, 201–218.

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Industrial and Labor Relations Review 43*(2), 245–257.

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. *Journal of Labor Economics 19*(1), 22–64.

Card, D. and J. DiNardo (2000). Do Immigrant Inflows Lead to Native Outflows? *American Economic Review*, 360–367.

Carrington, W. J. and P. J. F. de Lima (1996). The Impact of 1970s Repatriates from Africa on the Portuguese Labor Market. *Industrial and Labor Relations Review 49*(2), 330–347.

Chesher, A. and I. Jewitt (1987). The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator. *Econometrica: Journal of the Econometric Society*, 1217–1222.

Conley, T. (1999). GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics 92*(1), 1–45.

Conley, T., C. Hansen, and P. Rossi (2008). Plausibly Exogenous. Mimeo. Graduate School of Business, University of Chicago.

Conley, T. G. (2008). Spatial Econometrics. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics 19*(1), 15–18.

Davis, D. and D. Weinstein (2002). Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review 92*(5), 1269–1289.

Davis, D. and D. Weinstein (2008). Search for Multiple Equilibria in Urban Industrial Structure. *Journal of Regional Science 48*(1), 29–65.

Ellison, G. and E. Glaeser (1997). Geographic concentration in US manufacturing industries: a dartboard approach. *Journal of political economy 105*(5), 889–927.

Ellison, G., E. Glaeser, and W. Kerr (2009). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review, forthcoming*.

Filer, R. (1992). The Effect of Immigrant Arrivals on Migratory Patterns of Native Workers. In G. J. Borjas and R. B. Freeman (Eds.), *Immigration and the Workforce: Economic Consequences for the United States and Source Areas*, pp. 245–69. University of Chicago Press.

Frey, W. (1995). Immigration and Internal Migration 'Flight' from US metropolitan Areas: Toward a New Demographic Balkanisation. *Urban Studies 32*(4-5), 733–758.

Friedberg, R. (2001). The Impact of Mass Migration on the Israeli Labor Market. *Quarterly Journal of Economics 116*(4), 1373–1408.

Gandal, N., G. Hanson, and M. Slaughter (2004). Technology, Trade, and Adjustment to Immigration in Israel. *European Economic Review 48*(2), 403–428.

Greenstone, M., R. Hornbeck, and E. Moretti (2009). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy, forthcoming*.

Hanson, G. (2005). Market Potential, Increasing Returns and Geographic Concentration. *Journal of International Economics 67*(1), 1–24.

Hanson, G. and M. Slaughter (2002). Labor-Market Adjustment in Open Economies: Evidence from US states. *Journal of International Economics 57*(1), 3–29.

Hanson, G. and C. Xiang (2004). The home-market effect and bilateral trade patterns. *American Economic Review*, 1108–1129.

Harris, C. (1954). The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers 44*(4), 315–348.

Head, K. and T. Mayer (2004a). Market Potential and the Location of Japanese Investment in the European Union. *Review of Economics and Statistics 86*(4), 959–972.

Head, K. and T. Mayer (2004b). The Empirics of Agglomeration and Trade. In J. V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4, Chapter 59, pp. 2609–2669. Elsevier.

Hunt, J. (1992). The Impact of the 1962 Repatriates from Algeria on the French Labor Market. *Industrial and Labor Relations Review 45*(3), 556–572.

Kirby, D. (2006). *A Concise History of Finland*. Cambridge University Press.

Kotavaara, O., H. Antikainen, and J. Rusanen (2010). Accessibility Concentrating Population: a GIS Analysis Relating the History of Finnish Urbanization to the Accessibility by Road and Railway Networks in 1880–1970. Mimeo. University of Oulu.

Krugman, P. (1991a). *Geography and Trade*. MIT Press.

Krugman, P. (1991b). Increasing Returns and Economic Geography. *Journal of Political Economy 99*(3), 483.

Laitinen, E. (1995). Vuoden 1945 maanhankintalain synty. In E. Laitinen (Ed.), *Rintamalta raiviolle. Sodanjälkeinen asutustoiminta 50 vuotta*, pp. 52–138. Jyväskylä: Atena Kustannus.

Lewis, E. (2003). Local, Open Economies Within the US: How Do Industries Respond to Immigration? Working Paper 04-01. Federal Reserve Bank of Philadelphia.

Lewis, E. (2005). Immigration, Skill Mix, and the Choice of Technique. Working Paper 05-08. Federal Reserve Bank of Philadelphia.

Miguel, E. and G. Roland (2006). The Long Run Impact of Bombing Vietnam. NBER Working Paper No. W11954.

Ngai, L. and C. Pissarides (2007). Structural Change in a Multisector Model of Growth. *American Economic Review 97*(1), 429–443.

North, D. C. and R. P. Thomas (1973). *The Rise of the Western World. A New Economic History*. Cambridge University Press.

Overman, H., S. Redding, and A. Venables (2003). The Economic Geography of Trade, Production, and Income: A Survey of Empirics. In J. Harrigan and E. Kwan Choi (Eds.), *Handbook of International Trade*. Oxford: Blackwell.

Pihkala, K. (1952). The Land Settlement Program of Finland. *Land Economics 28*(2), 147–159.

Redding, S. and D. Sturm (2008). The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review 95*, 1766–1797.

Redding, S. J., D. M. Sturm, and N. Wolf (2007). History and Industry Location: Evidence from German Airports. CEPR Discussion Papers 6345.

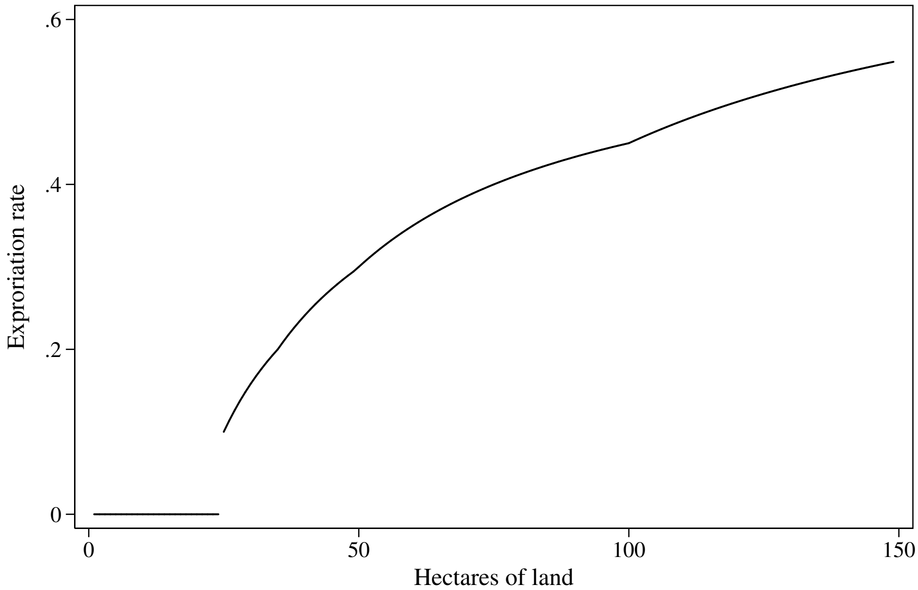
Rybczynski, T. (1955). Factor Endowments and Relative Commodity Prices. *Economica 22*(88), 336–341.

Sarvimäki, M., R. Uusitalo, and M. Jäntti (2009). Long-Term Effects of Forced Migration. SERC Discussion Paper 15.

Waris, H., V. Jyrkilä, K. Raitasuo, and J. Siipi (1952). *Siirtoväen sopeutuminen*. Otava.

Wright, R., M. Ellis, and M. Reibel (1997). The Linkage Between Immigration and Internal Migration in Large Metropolitan Areas in the United States. *Economic Geography*, 234–254.

Figure 1: Expropriation Rate for Privately Owned Agricultural Land



Note: The scale for land expropriation for private land owners. Set by Resolution of the Council of State in June 1945 and amended in July 1946. The size of the farm was determined on a basis of the total area of cultivated land, cultivable meadow and open pasture land. Farmers with two or more dependent children received some exemptions. Source: Pihkala (1952, Table II).

Figure 2: Spatial Distribution of the Instruments and the Displaced Persons

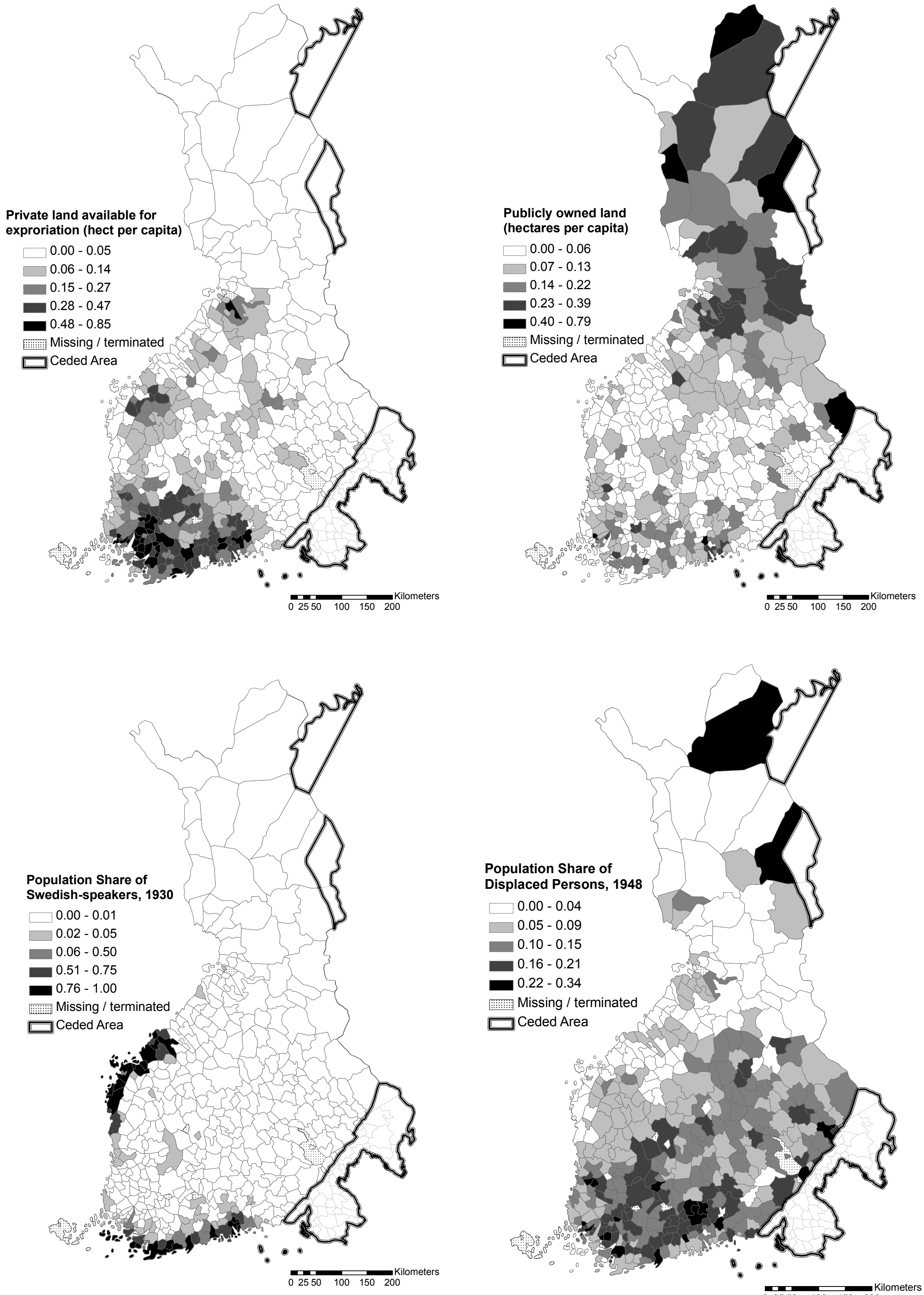
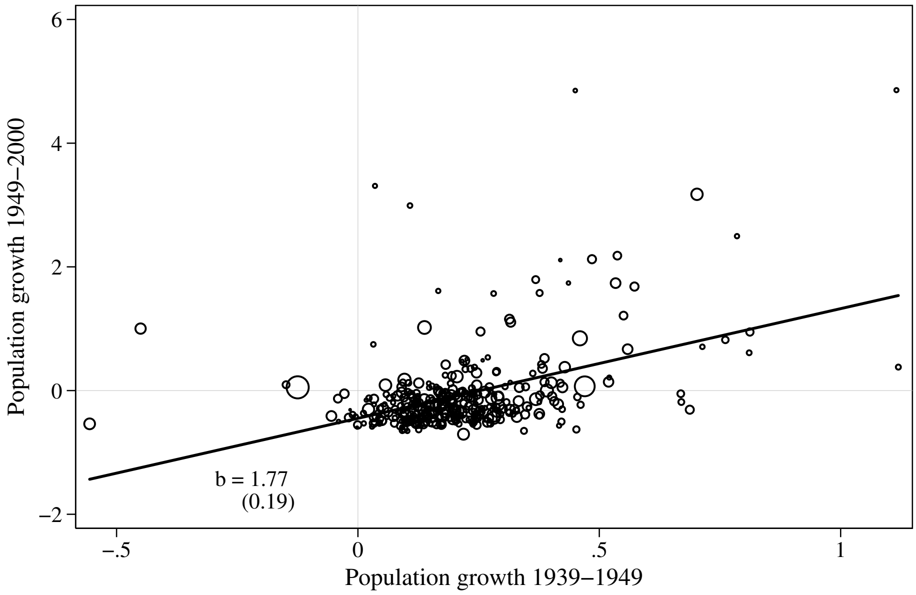


Figure 3: wartime and Post-War Population Growth



Note: Scatter plot and fitted values from regressing growth rate in 1949–2000 on growth rate in 1939–1949. Size of the dots correspond to the 1939 population.

Table 1: Impact of the Instruments on Pre-War and wartime Population Growth

	Falsification Test (pop. growth 1930–1939)		First-Stage (pop. growth 1939–1949)	
	(1)	(2)	(3)	(4)
Hectares of expropriable land per capita (1930)	-0.18***	-0.04	0.24***	0.22***
	(0.07)	(0.08)	(0.05)	(0.06)
Hectares of publicly owned land per capita (1940)	0.00	-0.14	0.25**	0.21**
	(0.14)	(0.15)	(0.10)	(0.10)
Share of Swedish-speaking population (1930)	-0.04	-0.06	-0.19***	-0.18***
	(0.05)	(0.05)	(0.04)	(0.04)
Control variables	no	yes	no	yes
F-statistic for instruments	2.8	0.9	21.5	22.6
Partial R^2	0.02	0.01	0.16	0.17

Note: The Coefficients for the instruments and standard errors (in parentheses). Sample: 349 rural municipalities. Control variables for column 2: share of labor force in primary sector 1930, population density in 1930, indicator for being a neighbor of a city (pre-war definition), longitude, latitude and nominal market access in 1930 (see footnote 11). Control variables for column 4: population growth between 1930 and 1939, mean taxable income per capita in 1938, share of labor force in primary sector 1930, population density in 1939, indicator for being a neighbor of a city (pre-war definition), longitude, latitude, nominal market access in 1939 and an indicator for being connected to railway network in 1939. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table 2: Impact of wartime Population Growth on Post-War Growth

	Dependent variable: Population Growth between 1949 and					
	1950	1960	1970	1980	1990	2000
	(1)	(2)	(3)	(4)	(5)	(6)
A: Baseline						
OLS	0.05***	0.52***	0.92***	1.31***	1.57***	1.77***
	(0.01)	(0.05)	(0.08)	(0.13)	(0.16)	(0.19)
2SLS	0.08***	0.23*	0.43**	0.66**	0.98**	1.28***
	(0.02)	(0.14)	(0.21)	(0.33)	(0.41)	(0.49)
B: Controlling for pre-war municipality characteristics and geography						
OLS	0.05***	0.47***	0.81***	1.10***	1.33***	1.48***
	(0.01)	(0.06)	(0.08)	(0.13)	(0.16)	(0.20)
2SLS	0.11***	0.45***	0.65***	0.89***	1.23***	1.48***
	(0.02)	(0.13)	(0.19)	(0.30)	(0.39)	(0.47)

Note: Coefficients for the population growth between 1939 and 1949 and standard errors (in parentheses). Instruments: Per capita privately owned agricultural land available for expropriation using 1930 farm size distribution, per capita government owned agricultural land in 1940, share of Swedish-speaking population in 1930. Sample: 349 rural municipalities. Control variables: population growth between 1930 and 1939, taxable income per capita in 1939, share of labor force in primary sector 1930, population density in 1939, indicator for being a neighbor of a city (pre-war definition), longitude, latitude, nominal market access in 1939 and an indicator for being connected to railway network in 1939. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table 3: Impact on Post-War Migration, Fertility and Mortality

	Net Migration	Internal Migration In	Internal Migration Out	Net Emigration	Fertility	Mortality
	(1)	(2)	(3)	(4)	(5)	(6)
A: Baseline						
OLS	0.92***	3.72***	2.82***	-0.02**	0.61***	0.10***
	(0.12)	(0.32)	(0.22)	(0.01)	(0.08)	(0.02)
2SLS	0.67**	3.58***	3.09***	-0.18***	0.12	0.00
	(0.30)	(0.80)	(0.55)	(0.03)	(0.20)	(0.05)
B: Controlling for pre-war municipality characteristics and geography						
OLS	0.69***	3.09***	2.41***	-0.01*	0.55***	0.07***
	(0.11)	(0.31)	(0.22)	(0.01)	(0.07)	(0.02)
2SLS	0.47*	3.82***	3.50***	-0.15***	0.50***	-0.04
	(0.27)	(0.76)	(0.55)	(0.03)	(0.16)	(0.04)

Note: Coefficients for the population growth between 1939 and 1949 and standard errors (in parentheses). Dependent variables are $\sum_{1951}^{T} m_{jt}/p_{j1951}$, where m_{jt} is the outcome variable and p_{j1951} is municipality's population in 1951. Instruments and control variables: see Table 2. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table 4: Impact of wartime Population Growth on Post-War Production Structure

	Dependent variable: Growth of Labor Force between 1950 and				
	1960	1970	1980	1990	2000
A: Primary Sector					
OLS	0.01***	0.02***	0.02**	0.03***	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2SLS	0.02*	0.03	-0.01	0.00	-0.01
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)
B: Non-Primary Sector					
OLS	0.06***	0.12***	0.19***	0.26***	0.23***
	(0.02)	(0.03)	(0.05)	(0.07)	(0.08)
2SLS	0.07**	0.16***	0.29**	0.55***	0.54***
	(0.04)	(0.06)	(0.11)	(0.17)	(0.20)

Note: Coefficients for the population growth between 1939 and 1949 and standard errors (in parentheses). Dependent variables: percentage change in the number of individuals working in primary (panel A) and non-primary (panel B) sector between 1950 and the year indicated by the columns. Instruments: see Table 5. Controlling for population growth between 1930 and 1939, taxable income per capita in 1939, share of labor force in primary sector 1930, population density in 1939, indicator for being a neighbor of a city (pre-war definition), longitude, latitude, nominal market access in 1939 and an indicator for being connected to railway network in 1939. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table 5: The Impact of Initial Conditions: Production Structure

	Dependent variable: Population Growth between 1949 and					
	1950	1960	1970	1980	1990	2000
A: OLS						
Population growth	0.07***	0.68***	1.11***	1.53***	1.85***	2.09***
between 1939 and 1949	(0.01)	(0.06)	(0.09)	(0.14)	(0.18)	(0.22)
Above Median LFS	0.01*	0.10***	0.12***	0.12*	0.12	0.11
in Agriculture, 1930	(0.00)	(0.03)	(0.04)	(0.06)	(0.08)	(0.10)
Interaction	-0.03*	-0.59***	-0.86***	-1.17***	-1.44***	-1.68***
	(0.02)	(0.10)	(0.15)	(0.24)	(0.31)	(0.38)
B: 2SLS						
Population growth	0.15***	0.60***	0.78***	1.00**	1.43***	1.69***
between 1939 and 1949	(0.03)	(0.17)	(0.24)	(0.39)	(0.50)	(0.61)
Above Median LFS	0.03**	0.07	0.02	-0.04	-0.04	-0.08
in Agriculture, 1930	(0.01)	(0.06)	(0.08)	(0.13)	(0.17)	(0.20)
Interaction	-0.10**	-0.45*	-0.42	-0.46	-0.71	-0.75
	(0.04)	(0.25)	(0.36)	(0.58)	(0.74)	(0.90)

Note: Coefficients for the population growth between 1939 and 1949, a dummy for municipality's labor force share in agriculture being above sample median in 1930 and their interactions. Each column in each panel comes from a separate regressions. Instruments: Per capita privately owned agricultural land available for expropriation using 1930 farm size distribution, per capita government owned agricultural land in 1940, share of Swedish-speaking population in 1930, all interacted with the agricultural LFS above median dummy. Controlling for population growth between 1930 and 1938, population density in 1938, indicator for being a neighbor of a city (pre-war definition), longitude, latitude, nominal market access in 1938 and being connected to the railway network in 1939. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table 6: The Impact of Initial Conditions: Railways

	Dependent variable: Population Growth between 1949 and					
	1950	1960	1970	1980	1990	2000
A: OLS						
Population growth between 1939 and 1949	0.04***	0.40***	0.60***	0.74***	0.91***	0.98***
	(0.02)	(0.09)	(0.13)	(0.21)	(0.27)	(0.33)
Connected to the Railway Network, 1939	-0.01*	-0.02	-0.06	-0.12*	-0.16*	-0.20**
	(0.00)	(0.03)	(0.04)	(0.06)	(0.08)	(0.10)
Interaction	0.02	0.10	0.31**	0.52**	0.61*	0.73*
	(0.02)	(0.11)	(0.16)	(0.25)	(0.32)	(0.39)
B: 2SLS						
Population growth between 1939 and 1949	0.08**	0.43**	0.54**	0.65	0.97*	1.19*
	(0.03)	(0.19)	(0.27)	(0.43)	(0.56)	(0.68)
Connected to the Railway Network, 1939	-0.03***	-0.01	-0.04	-0.12	-0.17	-0.21
	(0.01)	(0.06)	(0.08)	(0.13)	(0.16)	(0.20)
Interaction	0.07	0.07	0.26	0.55	0.62	0.68
	(0.04)	(0.25)	(0.36)	(0.56)	(0.73)	(0.88)

Note: Coefficients for the population growth between 1939 and 1949, a dummy for municipality being connected in the railway network in 1939 and their interactions. Each column in each panel comes from a separate regressions. Instruments: Per capita privately owned agricultural land available for expropriation using 1930 farm size distribution, per capita government owned agricultural land in 1940, share of Swedish-speaking population in 1930, all interacted with the agricultural LFS above median dummy. Controlling for population growth between 1930 and 1938, population density in 1938, indicator for being a neighbor of a city (pre-war definition), longitude, latitude and nominal market access in 1938 (see footnote 11). ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Figure A1: Alternative Sample Areas

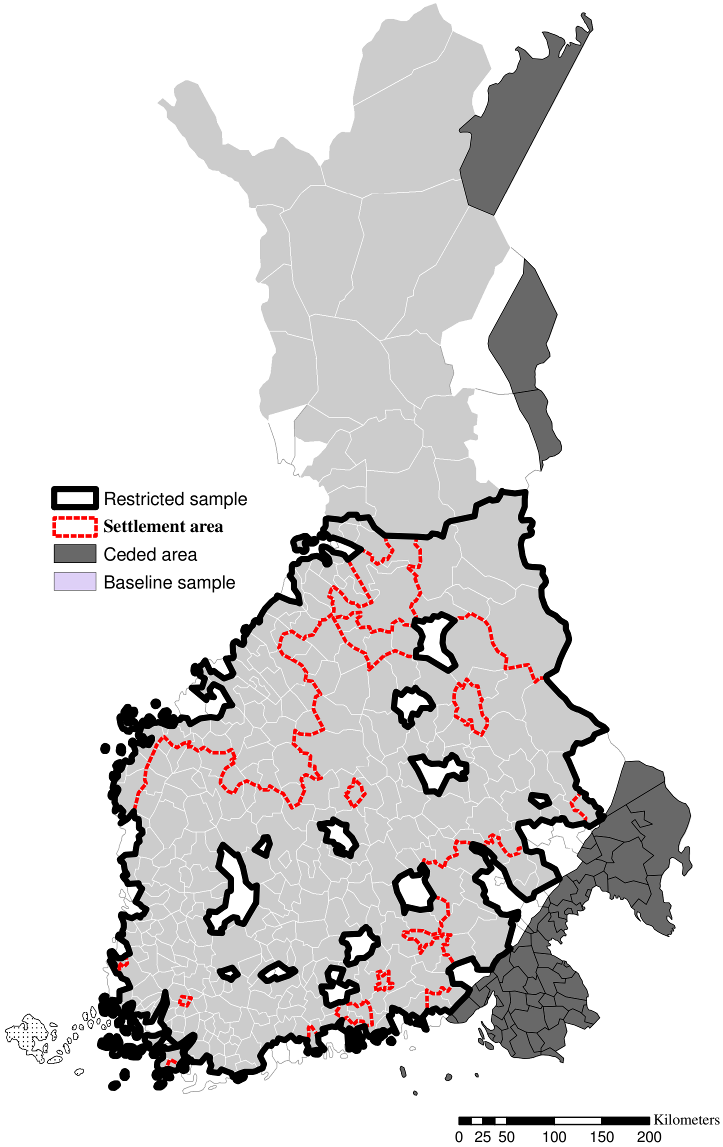


Table A1: Pre-War and wartime Population Growth by Instrument

	Falsification Exercise (pop. growth 1930–1939)			First-Stage (pop. growth 1939–1949)		
	(1)	(2)	(3)	(4)	(5)	(6)
A: Baseline						
Hectares of expropriable land per capita (1930)	-0.18***			0.26***		
	(0.07)			(0.05)		
Hectares of publicly owned land per capita (1940)		-0.08			0.38***	
		(0.14)			(0.11)	
Share of Swedish-speaking population (1930)			-0.04			-0.19***
			(0.05)			(0.04)
F-statistic for the instrument	7.6	0.3	0.7	28.8	12.7	24.2
Partial R^2	0.02	0.00	0.00	0.08	0.04	0.07
B: Controlling for pre-war municipality characteristics and geography						
Hectares of expropriable land per capita (1930)	-0.06			0.31***		
	(0.07)			(0.05)		
Hectares of publicly owned land per capita (1940)		-0.15			0.42***	
		(0.14)			(0.10)	
Share of Swedish-speaking population (1930)			-0.05			-0.22***
			(0.05)			(0.04)
F-statistic for the instrument	0.6	1.3	0.9	33.1	17.7	35.9
Partial R^2	0.00	0.00	0.00	0.09	0.05	0.10

Note: The impact of the instruments on the change in population growth between 1939 and 1949 and standard errors (in parentheses). Sample: 349 rural municipalities. Control variables for columns 1 to 3: share of labor force in primary sector 1930, population density in 1930, indicator for being a neighbor of a city (pre-war definition), longitude, latitude and nominal market access in 1930. Control variables for column 4 to 6: population growth between 1930 and 1938, mean taxable income per capita in 1938, share of labor force in primary sector 1930, population density in 1938, indicator for being a neighbor of a city (pre-war definition), longitude, latitude, nominal market access in 1939 and an indicator for being connected to railway network in 1939. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively

Table A2: Post-War Growth Estimates by Instrument

	Dependent variable: Change in log population share between 1949 and					
	1950	1960	1970	1980	1990	2000
	(1)	(2)	(3)	(4)	(5)	(6)
A: Baseline						
Hectares of expropriable	0.00	-0.45*	-0.02	0.57	1.18**	1.84***
land per capita (1930)	(0.03)	(0.27)	(0.34)	(0.47)	(0.59)	(0.70)
Hectares of publicly owned	0.17***	0.88***	1.07**	1.21*	1.59*	1.79*
land per capita (1940)	(0.05)	(0.30)	(0.42)	(0.67)	(0.86)	(1.03)
Share of Swedish-speaking	0.13***	0.74***	0.68**	0.55	0.56	0.49
population (1930)	(0.04)	(0.21)	(0.31)	(0.51)	(0.66)	(0.80)
All instruments	0.08***	0.23*	0.43**	0.66**	0.98**	1.28***
	(0.02)	(0.14)	(0.21)	(0.33)	(0.41)	(0.49)
Sargan test-statistic	10.5	27.0	7.2	0.8	0.9	2.2
(p-value)	0.00	0.00	0.05	0.66	0.61	0.40
B: Controlling for pre-war municipality characteristics and geography						
Hectares of expropriable	0.09***	0.19	0.48*	0.95**	1.52***	2.07***
land per capita (1930)	(0.03)	(0.19)	(0.27)	(0.41)	(0.54)	(0.66)
Hectares of publicly owned	0.12***	0.61**	1.00***	1.20**	1.59**	1.93**
land per capita (1940)	(0.04)	(0.25)	(0.35)	(0.55)	(0.72)	(0.87)
Share of Swedish-speaking	0.14***	0.61***	0.68***	0.75*	0.90*	0.88
population (1930)	(0.03)	(0.18)	(0.25)	(0.40)	(0.52)	(0.63)
All instruments	0.11***	0.45***	0.65***	0.89***	1.23***	1.48***
	(0.02)	(0.13)	(0.19)	(0.30)	(0.39)	(0.47)
Sargan test-statistic	2.0	5.5	2.3	0.3	0.6	1.9
(p-value)	0.39	0.08	0.33	0.78	0.61	0.34

Note: 2SLS estimates and standard errors (in parentheses) by instruments. The overidentification test reported at the last row is the p-value from of Sargan test-statistics. Sample: 349 rural municipalities. Control variables: see see Table 2. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table A3: Sensitivity to Violations of the Exclusion Restriction

		$\gamma_2 = 0$	$\gamma_2 = .1$	$\gamma_2 = .2$	$\gamma_2 = .3$
$\gamma_1 = 0$					
	$\gamma_3 = 0$	1.48***	1.41***	1.34***	1.27***
		(0.47)	(0.47)	(0.47)	(0.47)
	$\gamma_3 = -.1$	1.22**	1.15**	1.08**	1.01**
		(0.47)	(0.47)	(0.48)	(0.48)
	$\gamma_3 = -.2$	0.96**	0.89*	0.82*	0.75
		(0.48)	(0.48)	(0.48)	(0.48)
	$\gamma_3 = -.3$	0.70	0.63	0.56	0.49
		(0.49)	(0.49)	(0.49)	(0.49)
$\gamma_1 = .1$					
	$\gamma_3 = 0$	1.31***	1.24***	1.17**	1.10**
		(0.47)	(0.47)	(0.47)	(0.48)
	$\gamma_3 = -.1$	1.05**	0.98**	0.91*	0.84*
		(0.48)	(0.48)	(0.48)	(0.48)
	$\gamma_3 = -.2$	0.79	0.72	0.65	0.58
		(0.48)	(0.49)	(0.49)	(0.49)
	$\gamma_3 = -.3$	0.53	0.46	0.39	0.32
		(0.49)	(0.50)	(0.50)	(0.50)
$\gamma_1 = .2$					
	$\gamma_3 = 0$	1.14**	1.07**	1.00**	0.93*
		(0.47)	(0.47)	(0.47)	(0.48)
	$\gamma_3 = -.1$	0.88*	0.81*	0.74	0.67
		(0.48)	(0.48)	(0.48)	(0.48)
	$\gamma_3 = -.2$	0.62	0.55	0.48	0.41
		(0.48)	(0.49)	(0.49)	(0.49)
	$\gamma_3 = -.3$	0.36	0.29	0.22	0.15
		(0.49)	(0.50)	(0.50)	(0.50)
$\gamma_1 = .3$					
	$\gamma_3 = 0$	0.97**	0.90*	0.83*	0.76
		(0.48)	(0.48)	(0.48)	(0.48)
	$\gamma_3 = -.1$	0.71	0.64	0.57	0.50
		(0.48)	(0.48)	(0.48)	(0.49)
	$\gamma_3 = -.2$	0.45	0.38	0.31	0.24
		(0.49)	(0.49)	(0.49)	(0.49)
	$\gamma_3 = -.3$	0.19	0.12	0.05	-0.02
		(0.50)	(0.50)	(0.50)	(0.51)

Note: 2SLS estimates of α from estimation equation $(G_{t+1} - Z\gamma_0) = \alpha G_t + X\beta + u$. Parameters γ_1, γ_2 and γ_3 refer to the assumed direct effect of hectares of exproriable privately owned land per capita, hectares of government owned land per capita and the share of Swedish-speaking population in 1930, respectively. Outcome: Growth rate between 1949 and 2000. Control variables: see Table 2. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table A4: Sensitivity to Outliers

	Dependent variable: Population Growth between 1949 and					
	1950	1960	1970	1980	1990	2000
A: Baseline						
OLS	0.05***	0.47***	0.81***	1.10***	1.33***	1.48***
	(0.01)	(0.17)	(0.22)	(0.31)	(0.38)	(0.43)
2SLS	0.11**	0.45***	0.65***	0.89***	1.23***	1.48***
	(0.05)	(0.11)	(0.16)	(0.23)	(0.30)	(0.36)
Observations	349	349	349	349	349	349
B: Excluding observations with $D > .0236$						
OLS	0.03***	0.32***	0.58***	0.74***	1.03***	1.16***
	(0.01)	(0.05)	(0.08)	(0.12)	(0.17)	(0.20)
2SLS	0.06***	0.33***	0.57***	0.65***	0.97***	1.16***
	(0.01)	(0.10)	(0.15)	(0.23)	(0.31)	(0.38)
Observations	341	335	334	332	333	333
C: Excluding observations with $D > .0118$						
OLS	0.03***	0.32***	0.56***	0.76***	0.91***	1.00***
	(0.01)	(0.05)	(0.08)	(0.12)	(0.14)	(0.16)
2SLS	0.06***	0.30***	0.51***	0.56**	0.64**	0.83**
	(0.01)	(0.11)	(0.13)	(0.23)	(0.28)	(0.34)
Observations	338	327	322	322	323	326
D: Excluding observations with $D > .0059$						
OLS	0.03***	0.34***	0.53***	0.76***	0.97***	1.04***
	(0.01)	(0.04)	(0.07)	(0.12)	(0.14)	(0.16)
2SLS	0.05***	0.31***	0.46***	0.43**	0.65**	0.95***
	(0.01)	(0.09)	(0.14)	(0.20)	(0.28)	(0.31)
Observations	328	317	310	308	317	315
E: Excluding observations with $D > .0030$						
OLS	0.04***	0.34***	0.49***	0.69***	0.93***	0.85***
	(0.01)	(0.04)	(0.07)	(0.13)	(0.14)	(0.15)
2SLS	0.06***	0.32***	0.36***	0.38*	0.59**	0.70**
	(0.01)	(0.07)	(0.12)	(0.20)	(0.25)	(0.29)
Observations	304	301	290	288	298	296

Note: D refers to Cook's Distance measures (see the text for discussion). Instruments and control variables: see see Table 2. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table A5: Alternative Sample Areas

	Dependent variable: Population Growth between 1949 and					
	1950	1960	1970	1980	1990	2000
A: Baseline Sample (N=349)						
OLS	0.05***	0.47***	0.81***	1.10***	1.33***	1.48***
	(0.01)	(0.06)	(0.08)	(0.13)	(0.16)	(0.20)
2SLS	0.11***	0.45***	0.65***	0.89***	1.23***	1.48***
	(0.02)	(0.13)	(0.19)	(0.30)	(0.39)	(0.47)
B: Restricted Sample (N=330)						
OLS	0.04***	0.43***	0.75***	1.06***	1.29***	1.46***
	(0.01)	(0.06)	(0.09)	(0.14)	(0.18)	(0.22)
2SLS	0.10***	0.29*	0.30	0.56	0.96*	1.29**
	(0.03)	(0.17)	(0.25)	(0.38)	(0.49)	(0.59)
C: Settlement Area only (N=234)						
OLS	0.05***	0.48***	0.83***	1.18***	1.44***	1.62***
	(0.01)	(0.07)	(0.09)	(0.14)	(0.18)	(0.21)
2SLS	0.06***	0.25	0.31	0.52	0.98**	1.32**
	(0.02)	(0.19)	(0.27)	(0.39)	(0.49)	(0.56)
D: All (baseline sample augmented with cities, N=382)						
OLS	-0.03*	0.38***	0.68***	0.96***	1.23***	1.44***
	(0.01)	(0.05)	(0.09)	(0.13)	(0.17)	(0.21)
2SLS	0.07*	0.39***	0.56***	0.73**	1.02**	1.22**
	(0.04)	(0.13)	(0.21)	(0.32)	(0.41)	(0.50)

Note: Coefficients for the population growth between 1939 and 1949 and standard errors (in parentheses). Instruments and control variables: see Table 2, panel D also includes an indicator variables taking value one if the municipality was classified as a city before the war. ***, **, * indicate that the coefficient differs from zero at 1%, 5%, 10% level, respectively.

Table A6: Alternative Standard Errors

	Dependent variable: Population Growth between 1949 and					
	1950	1960	1970	1980	1990	2000
A: OLS						
Coefficient	0.052	0.473	0.810	1.100	1.327	1.481
Standard errors:						
Conventional	(0.009)	(0.056)	(0.080)	(0.125)	(0.163)	(0.197)
Robust	(0.013)	(0.167)	(0.218)	(0.306)	(0.379)	(0.434)
Spatial (50km)	(0.014)	(0.176)	(0.227)	(0.325)	(0.422)	(0.492)
Spatial (100km)	(0.014)	(0.175)	(0.221)	(0.314)	(0.411)	(0.481)
Spatial (250km)	(0.010)	(0.149)	(0.178)	(0.240)	(0.309)	(0.364)
B: 2SLS						
Coefficient	0.115	0.452	0.652	0.890	1.232	1.480
Standard errors:						
Conventional	(0.023)	(0.133)	(0.193)	(0.302)	(0.391)	(0.472)
Robust	(0.049)	(0.111)	(0.157)	(0.228)	(0.301)	(0.364)
C: Spatial GMM						
Cut-off: 50km	0.084	0.396	0.603	0.823	1.100	1.316
	(0.030)	(0.136)	(0.170)	(0.247)	(0.355)	(0.451)
Cut-off: 100km	0.085	0.408	0.604	0.822	1.130	1.388
	(0.029)	(0.136)	(0.171)	(0.236)	(0.339)	(0.442)
Cut-off: 250km	0.095	0.427	0.634	0.831	1.221	1.576
	(0.025)	(0.136)	(0.157)	(0.199)	(0.266)	(0.345)

Note: OLS, 2SLS and Conley's (1999) spatial GMM estimates for the population growth between 1939 and 1949 and standard errors (in parentheses). Instruments and control variables: see Table 2. Conventional standard errors are square roots of the diagonal elements of the covariance matrix estimator $\hat{V}_c = (X'X)^{-1}\left(\sum \hat{\epsilon}_i^2/N\right)$, where $\hat{\epsilon}_i = y_i - X_i'\hat{\beta}$ is the estimated regression residual. Robust standard errors are square roots of the diagonal elements of the covariance matrix estimator $\hat{V}_r = N\,(X'X)^{-1}\left(\sum X_i X_i' \hat{\epsilon}_i^2/N\right)(X'X)^{-1}$. Spatial standard errors are square roots of the diagonal elements of the covariance matrix estimator $\hat{V}_s = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} K_N\left(s_i, s_j\right) \cdot z_{si}\hat{\epsilon}_i z_{sj}^{'}\hat{\epsilon}_j$, where $K_N\left(s_i, s_j\right)$ is an uniform kernel taking value one if locations s_i and s_j are within a cut-off distance of each other and zero otherwise.