

IZA DP No. 4003

Long-Term Effects of Forced Migration

Matti Sarvimäki
Roope Uusitalo
Markus Jäntti

February 2009

Long-Term Effects of Forced Migration

Matti Sarvimäki

London School of Economics

Roope Uusitalo

*Government Institute for Economic Research (VATT),
IFAU and IZA*

Markus Jäntti

Åbo Akademi University

Discussion Paper No. 4003
February 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Long-Term Effects of Forced Migration^{*}

We study the long-term effects of human displacement using individual-level panel data on forced migrants and comparable non-migrants. After World War II, Finland ceded a tenth of its territory to the Soviet Union and resettled the entire population living in these areas in the remaining parts of the country. We find that displacement increased the long-term income of men, but had no effect on that of women. We attribute a large part of the effect to faster transition from traditional (rural) to modern (urban) occupations among the displaced.

JEL Classification: J60, O15, R23

Keywords: migration, displaced persons, regional labor markets

Corresponding author:

Matti Sarvimäki
Spatial Economics Research Centre
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
United Kingdom
E-mail: m.sarvimaki@lse.ac.uk

^{*} We thank Jérôme Adda, Josh Angrist, Christian Dustmann, Per-Anders Edin, Peter Fredriksson, Albrecht Glitz, Kristiina Huttunen, Victor Lavy, Guy Michaels, Tuomas Pekkarinen, Marie Schultz-Nielsen, Daniel Sturm, Olof Åslund and seminar participants at Fordham, HECER, LSE/CEP, NYU, UCL/CReAM, Uppsala, VATT, European Society for Population Economics (Chicago), Econometric Society (Budapest), Nordic Summer Institute in Labor Economics (Helsinki), Society of Labor Economists (New York), Nordic Migration Workshop (Helsinki) and Finnish Society for Economic Research (Jyväskylä) for helpful comments and Tiina Hytönen, Nina Intonen and Anni-Mari Karvinen for able research assistance. Financial support from Yrjö Jahnsson Foundation is gratefully acknowledged. Sarvimäki also acknowledges financial support from Jenny and Antti Wihuri Foundation.

1 Introduction

Armed conflicts, natural disasters and infrastructure projects force people to migrate. While international refugee flows have dominated the attention of the Western world, most forced migrants are displaced within their home countries. According to the UNHCR's (2008) estimate, more than three quarters of the 67 million people forced to move at the end of 2007 were internally displaced. Sadly, forced migration is not likely to cease. If anything, new causes, such as global warming, may increase the number of displaced persons. For instance, a one meter rise in sea levels would permanently inundate the land currently hosting 11 percent of the Bangladeshi population (Agrawala et al., 2003). Even if such scenarios are far fetched, ordinary peace-time public policies sometimes lead to large-scale forced migrations. According to the World Commission on Dams (2000), for example, dam construction alone has displaced between 40 and 80 million people in the past half a century.

These displacements are likely to have important consequences for those who are forced to migrate as well as the receiving and the sending areas. They also create a major policy challenge. Understanding the consequences of human displacement and the mechanisms leading to these consequences would be essential for developing effective programs to assist people uprooted from their homes. Yet, the topic has been subject to relatively little economic research.¹

This paper studies the long-term effects of being internally displaced in Finland after the World War II. The war led Finland to cede a tenth of its territory to the Soviet Union. The entire population from 60 rural municipalities and three cities was evacuated within a few weeks. Altogether 430,000 individuals, 11 percent of the 1940 population, were resettled to the remaining parts of the country. We focus on the long-term impact of migration on those who were forced to migrate. To perform the analysis, we have access to unusual individual-level longitudinal data on the displaced

¹In his survey, the only papers Lucas (1997) finds on displaced persons are those by Schultz (1971) and Gottschang (1987). More recent research include Czaika and Kis-Katos (2007), Ibáñez and Vélez (2008) and Kondylis (2007, 2008).

and non-displaced populations from strictly comparable sources. Our data record the situation of more than 20,000 individuals before the war as well as their post-war outcomes up to fifty years after the displacement.

We find that forced migration increased long-term income. Before the war, those living in the areas that were to be ceded did not differ from the rest of the population. By contrast, they were substantially more mobile in the post-war period and earned significantly more in 1971. These findings illustrate that successful resettlement policies are possible. Moreover, they suggest that increased mobility had a positive effect on the long-term incomes of the displaced.

While displacement is an important topic on its own right, we argue that these findings also contribute to a more general understanding of migration. Economists typically consider migration as an equilibrating mechanism promoting efficient resource allocation. In this framework, migration is a human capital investment that involves both costs and benefits for the migrants (Sjaastad, 1962). In the absence of costs, labor flows would equalize the value of the marginal product of workers across labor markets and the output of the economy would be maximized. However, large regional wage differentials persist even after conditioning for observable characteristics of the residents.

The situation we analyze provides an interesting natural experiment that allows us to assess whether differences in unobservable characteristics or high migration costs give rise to regional disparities. That is, in equilibrium, the returns to migration are equal to its costs for a marginal migrant. Therefore, a consistent estimate of returns to migration would be informative about the magnitude of these costs. The challenge to estimation is that if returns to migration vary across individuals, those who decide to migrate are a self-selected group of the population. As a consequence, correlations between migrant status and outcomes of interest in a typical observational dataset are likely to overstate the returns to migration (see Greenwood, 1997, for a survey). In our case, the selection problem does not occur, since the entire population in a certain area was forced to migrate. Therefore, we estimate returns to (forced) migration for an essentially randomly chosen individual

in the sample.

Our results suggest that large economic gains from migrating would have been available also for those who were not displaced. However, many chose not to exploit these opportunities. This is consistent with migration carrying a high cost. In essence, being displaced reduced these costs. This finding is highly relevant for regional policy. While we certainly do not advocate policies forcing people to migrate, the results suggest that less brutal policies, such as providing assistance for voluntary movers, may be more efficient than policies supporting those who remain in the less-advantaged regions. Furthermore, higher mobility would also be likely to help those who stay put by reducing regional and occupational wage differences (Borjas, 2001).

The rest of this paper is organized as follows. Section 2 provides the details of the post-WWII situation in Finland and the implemented settlement policy. Section 3 describes the data. Section 4 discusses our empirical approach. Section 5 present the results and section 6 provides robustness checks. We discuss the implications of the results in section 7. Section 8 concludes.

2 Historical Background

In August 1939, just before the outbreak of the World War II, Germany and the Soviet Union signed a non-aggression pact that included a secret clause in which the two signatories divided up Eastern Europe between themselves. Finland was consigned to the Soviet sphere of influence. Following unsuccessful negotiations about Soviet Union's territorial demands, the Red Army attacked Finland in November 1939, three months after the German invasion into Poland. In the peace treaty ending the battles between Finland and Soviet Union in March 1940, Finland ceded roughly a tenth of its territory to the Soviet Union. The entire population of these areas had been evacuated during the war. An Emergency Settlement Act (*Pika-asutuslaki*) was enacted in July 1940 to settle the displaced into the rest of the country.

The execution of the Act was suspended in June 1941, when Finland joined Germany's attack to the Soviet Union. By the end of August, the

Finnish troops had reoccupied the ceded areas and on December 6th, the Finnish Parliament declared them re-united with the rest of Finland. Approximately two thirds of the displaced returned to their pre-war homes.

However, in the summer of 1944, the Red Army pushed the Finnish troops back to roughly the same line of defense they had held at the end of the first war. The entire population of the ceded areas was again evacuated. The armistice signed in September 1944, and later ratified in the Paris peace treaty in 1947, restored the 1940 borders with some additional areas ceded to the Soviet Union. Finland also agreed to pay USD300 million in war reparations and to expel German troops from its territory, which led to Finland's third armed conflict during World War II.

Figure 1 presents the ceded areas on a 1938 map of Finland.² Most of the displaced lived in the region of Karelia in the South-East of the country, while the ceded areas in the North were extremely sparsely populated. The map also displays average incomes, illustrating that the ceded area include both rich and poor areas. As we will show below, the average incomes in the ceded area was close to the national average.

The war left Finland with approximately 92,000 dead and 228,000 injured out of total population of four million. Much of the country's production capacity was destroyed in the war and further cuts in capacity were caused by war reparations. For example, a quarter of the Finnish commercial fleet was handed over to the Soviet Union. Altogether war reparations took about 15 percent of the government budget between 1945 and 1949 (Tarkka, 1988).

Even without such a dire situation, settling the 430,000 displaced persons would have been a major burden. Finland was still a predominantly agrarian society, in many ways resembling current middle-income developing countries, with roughly half of the working age population employed in agriculture. Similarly, almost one half of those forced to migrate were farmers. The only feasible option at the time was to resettle a large fraction of the

²In addition to the area ceded in 1940, the Petsamo area, in the North of Finland, was ceded to the Soviet Union in 1944. Furthermore, the Porkkala Peninsula near Helsinki was leased for a Soviet naval base for fifty years. Following an improvement in international relations and changes in military technology that made land-based artillery less important for protecting Leningrad from the sea, Porkkala was returned to Finland in 1956.

displaced to areas where they could derive their main income from farming.

In May 1945, the Parliament approved the Land Acquisition Act (*Maanhankintalaki*), which guided the settlement policy. The displaced who had owned or rented land in the ceded areas and had received their principal income from agriculture were entitled to receive land from remaining parts of the country. Others received compensation for their lost property in the form of government bonds. Land was primarily taken from the state, the local governments (municipalities) and the church, but the required amount far exceeded the capacity of the public sector. Thus, roughly two thirds of the cultivated fields, one half of land that could be cleared for cultivation and a third of forest land was seized from private owners. The committee drafting the law proposed an explicit progressive expropriation schedule for seizing private land. Similarly, a progressive tax was set on other forms of property.³

The implementation of the Land Acquisition Act was entrusted to the Department of Land Settlement in the Ministry of Agriculture. Altogether 147 local land redemption boards were responsible for the expropriation measures and the same number of local settlement boards had a duty to locate applicants for land. Those coming from each Karelian village were settled into a designated target municipality. The number of displaced allocated in each municipality was mainly affected by the availability of suitable land, which again mainly depended on the pre-war farm size distribution and on the quantity of state-owned land in the municipality. The most important factor in allocation across receiving municipalities was the location of their municipality in the ceded area. Those from the western parts of the Karelian peninsula were settled along the southern coast, those from the eastern part of the Karelian peninsula north of the first group and those from Northern

³The schedule for farm land required private land owners to cede up to 80% of their land holdings depending on the size of their farms. No land was expropriated from farms smaller than 25 hectares. The landowners were compensated with government bonds yielding 4% nominal interest. Inflation eventually wiped out about four fifths of their value. However, the bonds could be used for paying the Property Expatriation Tax, which was collected from all capital owners in order to cover the costs of the resettlement. Pihkala (1952) discusses the land acquisition policy in detail and argues that landowners probably did not suffer more than other owners of property. Waris et al. (1952) provides details of the Property Expatriation Tax.

Karelia even further north. None were placed in Northern Finland, where conditions for agriculture are unfavorable. Those from the municipalities surrounding Viipuri, the largest city in Karelia, were settled close to the capital, Helsinki, and those from Sortavala, the second largest Karelian city, close to Jyväskylä, a city located in the middle part of the country. Persons from the same village were settled in the same municipality and neighboring villages were settled close to each other.

The destination of the non-agrarian displaced was mainly determined by the availability of housing and the distance from the ceded areas. Cities in Eastern Finland received flows of displaced persons that constituted almost ten percent of the population, while cities further west and cities with the most severe housing problems received much less.⁴ While the non-agrarian population was not explicitly allocated, the settlement plan influenced also their migration due to family ties and employment opportunities with their former landowner employers. In June 1949, 53 percent of the displaced lived in their designated placement areas (Waris et al., 1952).

3 Data

Our primary source of data is the 1950 population census, the first full census implemented in Finland. Data were collected by personal interviews and the information for each dwelling unit was stored on a single form. These forms were sorted by municipality, within municipalities in alphabetical order and then filed in boxes. In 1997, Statistics Finland drew a sample from the full 1950 census by picking every tenth box. Nearly all of the information on the census forms was then keyed into a database. The resulting sample contained about 114,000 dwelling units with 411,629 persons from 392 of Finland's 547 municipalities. Based on the first and the last names, along

⁴The share of the displaced in 1948 living in urban areas varied from 9.2% in Mikkeli, 8% in Jyväskylä and 7.8% in Lahti (all located in Central or Eastern Finland) to 2.6% in Pori (on the west coast) and 0.4% in Pietarsaari (a Swedish-speaking town on the western coast). Housing shortages in the capital city, Helsinki, 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.

with date and place of birth, Statistics Finland matched these data to the Population Register in order to find the social security number – which had been introduced in 1964 – for each person. Thus, in order to end up in the final sample, the person had to be alive and live in Finland in mid-1960s. Social security numbers allowed merging the 1950 census file to a longitudinal census file containing information from population censuses performed every five years starting in 1970. Hence, the 1950 census sample can be followed through each subsequent census up to the year 2000.⁵ Statistics Finland (1996) provides a detailed discussion of the data.

The 1950 census contains information on various household characteristics and person-level information such as place of birth and residence, education, occupation and sector of employment. Importantly, it also contains retrospective information concerning municipality of residence in 1939. Other information from 1939 includes occupational status and industry codes referring to September 1st, 1939 – two months before the war began. The same information is available for 1950. This creates an unusual situation where we have longitudinal microdata on the displaced and the non-displaced from pre- and post-displacement periods. The same survey instruments were administered to both groups and hence all information is fully comparable.

Linking the 1950 census to the longitudinal census data for 1970–2000 further increases the amount of available information. Most importantly, the 1970 census include tax record data from 1971. This provides an opportunity to evaluate the long-term effects of migration 27 years after the war ended. We are particularly interested in the effect of forced migration on individual productivity or productive capacity, measured by several labour market-related measures. Primary among those is taxable income. Even longer-term effects on income can be evaluated by using data on taxable income around 1980–1990 when most of the (surviving) displaced are already retired from the labor force. Since pension income is affected by the accrued pension rights from each employment spell during the career, the taxable income of

⁵Ideally we would also have data from the 1960 census. However the original punch cards were destroyed in the 1970s and the magnetic tapes (where the data had been stored) were damaged in storage.

pensioners is a reasonable proxy for lifetime income.

The most important shortcoming of the data is that the 1950 census did not collect any direct information on income. However, the data contain several variables that are informative about the economic status of the individuals. In order to summarize this information efficiently, we construct two measures of imputed income (see the Appendix for detail). Our main approach is to use the information available in the microdata on the 1971 taxable income. That is, we break down the data into 590 industry–socioeconomic status–living in an urban area categories using the 1939 and 1950 data and assign each observation the average 1971 log annual income of the respective cell. To complement this measure, we also use detailed annual tables listing taxable income based on reports from local tax boards (Statistics Finland, 1942, 1953). These data allow us to classify the observations into 38 industry-occupation-socioeconomic status groups for 1950 and 12 industry-living in urban area groups for 1939 and assign each individual the mean income of their reference group.

We have access to a smaller random sample of the data originally stored by Statistics Finland.⁶ In order to focus on those who were of working age throughout the period from 1939 to 1971, we further limit the analysis to individuals born between 1907 and 1925. Moreover, we exclude observations where municipality of residence in 1939 is unknown (312 persons), who lived in the 12 partly ceded municipalities (642 persons) or on the Åland Islands (178 persons) as well as those who lived in the 13 municipalities for which pre-war data on taxable income per capita is missing (498 persons). This leaves us with information on 22,771 individuals of whom 2,558 were displaced.

⁶Data used in the analysis contain confidential information from tax registers. All datasets and code used in the paper and their English language descriptions are available from the authors for replication purposes, but data access requires a prior approval by Statistics Finland. Details on data access policy and application procedure can be found from the Statistics Finland website at http://www.stat.fi/meta/tietosuoja/kayttolupa_en.html.

3.1 Descriptive Statistics

Table 1 reports sample means for the displaced and the rest of the population in 1939, 1950 and 1970/71. The pre-war means are relatively similar. The main difference is the lower share of blue-collar workers among the future displaced. This reflects partly the lower share working in manufacturing and partly the slightly higher share of assisting family members and those out of the labor force. Furthermore, the share working in the formal labor market, i.e. working as an entrepreneur or hired labor, and the share living in urban areas are lower among those who were later displaced. However, these differences are small in magnitude. Importantly, as we will see in more detail in Section 5, there are no substantial income differences by either definition of imputed income or by average taxable income per capita in the municipality of residence. The only large difference is that very few of those living in the ceded areas spoke Swedish as their mother tongue.

The next two columns report means in 1950. The most important changes are decreases in the share of the population employed in agriculture, by 17 and 9 percentage points among the displaced and the rest of the population, respectively. The shares of population working in manufacturing are now equal for the displaced and the others. The share working in service or other sectors increased among the displaced, but remained constant among the rest of the population. The fast flow away from agriculture among the displaced is also evident in that they are now more likely to work as blue-collar workers than the non-displaced. As a result, their imputed incomes are roughly six percent higher than those of the non-displaced. There are no large differences in educational attainment between the displaced and the non-displaced. Furthermore, the displaced were more likely to live in cities or market towns, though the difference remains small in magnitude. Not surprisingly – given that the displaced had lost their homes only six years earlier – there was a clear difference in the fraction living in owner-occupied housing.

The remaining columns report means in 1970/71. The share of the sample working in agriculture continues to decrease, but the gap between the

displaced and the rest of the population remains at ten percentage points. The displaced also remain more likely to work in the service and construction sectors, and now also in manufacturing. Interestingly, the difference in the proportion living in urban areas has increased markedly in comparison to 1950. As a consequence, the displaced now live in municipalities with higher mean taxable income. They are still less likely to live in owner-occupied housing. Most importantly, the displaced have higher annual incomes than the non-displaced, suggesting that forced migration or perhaps higher post-war mobility may have had a positive long-term effect on income.

4 Empirical Approach

We will evaluate the impact of being displaced by comparing the outcomes of those living in the ceded area to comparable persons who were not forced to move due to the war. Our base estimation equation is

$$y_{ijt} = \alpha D_i + \mathbf{X}_{0i}\beta + \varepsilon_{ijt} \quad (1)$$

where y_{ijt} is the outcome of interest for individual i living in location j at time t , D_i is a dummy indicating displacement status, \mathbf{X}_{0i} is a vector of observable characteristics measured before the war and ε_{ijt} summarizes the unobservable factors. In order to interpret the results and to state our identifying assumption clearly, we divide the unobservables into two parts

$$\varepsilon_{ijt} = u_{jt} + \nu_{it} \quad (2)$$

where u_{jt} captures the unobserved “quality” of the labor market j at time t and ν_{it} is individual specific error term.

Our identifying assumption is that, once we condition for the observed characteristics, displacement status is uncorrelated with unobserved individual characteristics:

$$Cov(D_i, \nu_{it} | \mathbf{X}_{0i}) = 0. \quad (3)$$

Since the location of the new border was determined as an outcome of the

battles, this assumption seems plausible. Those who happened to live on the ceded area had no option but to move. Nevertheless, one could argue that those living in the Western part of the country prior to the war could have differed from those living in the Eastern part. Below we provide two types of evidence suggesting that this was not the case. First, the available data suggest there were few important pre-war differences between the future displaced and non-displaced persons. Second, the estimates are not sensitive to controlling for pre-war observables characteristics or for excluding the Western part of the country from the analysis. Thus, we argue that assumption (3) holds and our estimates can be considered causal in a sense discussed next.

4.1 Interpretation of the Base Estimates

The key to interpreting our main results is to note that the correlation between displacement status (D_i) and post-war labor market quality (u_{jt}) is positive. This occurs for at least three reasons. First, resettlement moved individuals to new labor markets and occupations. While the displaced had limited possibilities to choose their initial destination, the authorities assigned more displaced persons to more prosperous municipalities. Second, the displacement appears to have hastened the transition from traditional to modern occupations. Third, the likelihood of moving again between geographic areas after the resettlement appears to have been higher among the displaced than among the rest of the population. These later moves are likely to be correlated with job opportunities. If the displaced had lower costs of moving, perhaps because they had less reason to stay in the placement area, post-war sorting across municipalities may differ between the displaced and the rest of the population.

These considerations can be summarized formally by noting that a least-squares estimator of α has the probability limit

$$\text{plim } \hat{\alpha} = \pi_D + \theta \tag{4}$$

where π_D is the partial correlation between displacement status and labor

market quality – where labor market refers to the interaction between a spatial location and an occupation – and parameter θ captures direct effects of being displaced.⁷ That is, the impact of forced migration works through two mechanisms. The first is moving to better labor markets. The second are “other” effects due to e.g. loss of location specific human capital, trauma or loss of property. The same interpretation carries over to quantile regressions.

Given assumption (3), $\hat{\alpha}$ has a causal interpretation in the sense that it measures the difference in expected outcomes between two well-defined counterfactual states. However, it is important to acknowledge that large-scale migration flows probably affected not only the displaced persons but the entire population of the post-war Finland. Thus the counterfactuals are (i) being displaced in post-war Finland and (ii) not being displaced in post-war Finland. In other words, one should think of a thought experiment where the displacement status of a single individual is manipulated, while 430,000 others are still forced to migrate.

4.2 Controlling for Current Labor Market Quality

Below we will also report estimates from specifications adding post-war labor market fixed-effects to equation (1). The motivation is to estimate θ by conditioning on current labor market characteristics. This would allow us to divide the impact of being displaced into a part due to increase in labor market quality and into a part due to other reasons. However, adding labor market fixed effects to equation (1) may not produce consistent estimate of θ . This is due to the fact that even if the displacement status D_i and individual specific factors ν_{it} are uncorrelated, they are uncorrelated *conditional on* u_{jt} only if $Cov(D_i, u_{jt}|\mathbf{X}) = 0$ or $Cov(u_{jt}, \nu_{it}|\mathbf{X}) = 0$. We have already argued that the first equality does not hold. The second equality would imply that individuals do not sort into localities based on their unobservable characteristics. This is also unlikely to be true. For instance, if those with

⁷To derive this probability limit, suppose that the data generating process is $y_{ijt} = \theta D_i + \mathbf{X}_{\mathbf{0}i}\beta + u_{jt} + \nu_{it}$. Equation (4) then follows from the familiar omitted variables bias, i.e. π_D is the probability limit of the OLS estimator of the displacement status in a regression of u_{jt} on D_i and $\mathbf{X}_{\mathbf{0}i}$.

above average unobserved characteristics are more likely to be located in above average labor markets, our estimate of θ will be biased.⁸ Even in this case, estimates controlling for labor market fixed effects contain useful information. They are the average differences between a displaced and a non-displaced person, who were similar in their pre-war characteristics and who lived in similar labor markets in 1970.

5 Results

We now report the results from regression models discussed above. In the following two Sections, we focus on summarizing the data and postpone a more speculative discussion on the implications of the results to section 7. We begin by asking whether we can find any evidence that the displaced differed from the rest of the population prior to the war. We then present our main results on the causal effect of forced migration among the displaced on several long-term outcomes. This is followed by studying the impact of conditioning on post-war labor market fixed-effects and heterogeneity of the main effects as well as going through several robustness checks.

5.1 Pre-War Differences

In Section 3 we saw that the pre-war mean characteristics of the displaced were rather similar to those of the rest of the population. We now perform simple tests of the significance of the pre-war differences by regressing the available pre-war information on an indicator variable taking value one if the person was to become displaced after the war and zero otherwise. Table 2 reports the results separately for men (panel A) and women (panel B), from a bivariate regression (column 2) and from a specification controlling for age, longitude and latitude of the 1939 residence municipality and dummies for living in urban area and speaking Swedish as one's mother tongue in 1939 (column 3). The outcomes considered are a dummy for being employed as an entrepreneur or hired labor, two definitions of imputed income and taxable

⁸See Angrist and Pischke (2009, Section 3.2.3) for a detailed discussion on the bias induced by conditioning on variables that are affected by the treatment.

income per capita in the 1939 municipality of residence. Both imputed income measures vary only between industry, socioeconomic status and urban versus rural residence. Thus, the estimates are informative only on whether the future displaced were employed in occupations carrying above average income.

The results suggest that pre-war differences between future displaced and the rest of the population were small or nonexistent. Out of the 16 estimates, only two are statistically significant at the 10% level or lower. These significant estimates would imply that men who were later displaced were working in pre-war occupations with a slightly higher average wages and that women living in the ceded areas were less likely to be employed in the formal labor market than comparable women in other parts of the country. Note, however, that given a large number of regressions, one should expect to get statistically significant estimates occasionally even if the outcomes were randomly generated. Thus we conclude that, on balance, we find little evidence suggesting that the economic performance of the displaced differed from the rest of the population prior to the war. Furthermore, as our base specification controls for these pre-war observable characteristics, assumption (3) is very likely to hold.

5.2 Main Results

We next turn to post-war outcomes. Tables 3, 4, 5 and 6 present our main results. Each table has a similar structure, where column (2) reports the results from regressing several post-war outcomes on a binary indicator of the person being displaced. Each coefficient comes from a separate regression. Under the assumption that being displaced was random, these estimates correspond to the parameter $\hat{\alpha}$ in equation (4). That is, they are informative about the causal effect of being forced to migrate inside Finland after the World War II. Column (3) reports similar estimates from a specification controlling for age, imputed income in 1939, longitude and latitude of the 1939 municipality of residence, and indicator variables for being Swedish-speaking and living in urban area in 1939. The main impact of

conditioning on these pre-war observable characteristics is the improvement in precision, while point estimates are virtually unaffected. The remaining columns present results conditioning on post-war labor market fixed-effects. We return to the interpretation of these estimates in the next subsection.

Consider first the impact of forced migration on short-term outcomes, measured in 1950. The point estimates in bottom rows of Table 3 suggest that displaced men were roughly ten percentage points more likely to change their sector of employment than non-displaced men. The estimates for imputed incomes show that these flows took place towards occupations carrying roughly 10 per cent higher wages (recall that imputed incomes are constructed as the mean income in cells defined by industry-socioeconomic status-living in urban area). As we already saw above, this predominantly reflects more frequent transitions from agriculture to manufacturing and services among the displaced, while transitions between categories of socioeconomic status are similar among both groups. Interestingly, however, we find no effect on the propensity to work in the formal labor market.

The impact on urbanization is less clear. While the point estimates suggest that displacement increased the likelihood of moving to urban areas, the results are not statistically significant.⁹ Furthermore, the estimates presented in Table 4 suggest that displaced women were as likely to change their sector of employment as other women. However, this result is due to three out of four of sector changes between 1939 and 1950 among women being due to leaving the category “unknown”. A closer look reveals that in comparison to other women, the displaced are less likely to switch from “unknown” sector of employment to agriculture and more likely to services. These transitions are also captured by the impact of the displacement on imputed income, which is similar to the estimates for men.

Consider next the long-term effects on the 1970 situation presented in Tables 5 and 6. We now find a positive impact on the propensity to be

⁹Note, however, that our inference is rather conservative as the standard errors are clustered at 1939 residence municipality level. We chose this approach in order to take into account that persons coming from same areas might have been exposed to common shocks, in particular due to the settlement plan affecting individuals based on their pre-war residence municipality.

employed in the formal labor market. More importantly, the data also allow us to study the impact of displacement on actual annual taxable income in 1971. The estimates for men suggest that forced migration substantially increased long-term income relative to the non-displaced. The point estimate for income levels (including zeros) indicates a difference of 2,600 Finnish marks, which corresponds to roughly an 11 percent difference in annual income. The estimate for log income implies that the impact would have been even larger. As we will show below, the difference in the estimated mean effect when measuring income in levels or logs is mostly due to the effect being larger in the lower end of the income distribution. Using log income as an outcome measure places more weight on those in the lower end of the income distribution, where the effect is large, and therefore yields higher estimates. The impact on pension income suggests a roughly 11 percent increase in lifetime income. For women, the results are mixed. While the impact on pension income is similar to that for men, the 1971 income of displaced women did not differ from that of other women.

On balance, these results suggest that being forced to migrate after World War II in Finland had a sizable positive causal effect on lifetime income. We next turn to possible sources of this effect. One candidate explanation is the transformation from traditional to modern occupations. As we already saw, being displaced increased the pace of this process. However, while the displacement had a permanent impact on the occupational distribution, the largest changes occurred soon after the war. For the period between 1950 and 1970, we find no difference in the likelihood of changing sector of employment. With regard to mobility on 1-digit occupational category level, we find a small but statistically significant positive effect for women. Nevertheless, the difference in imputed income remains virtually unchanged in comparison to short term effects.

However, estimates presented in the bottom rows of Tables 5 and 6 show that the displaced remained more mobile also in the period between 1950 and 1970. Instead of changing occupation, this mobility now takes the form of a higher propensity to change municipality of residence. Furthermore, among those who migrated, the displaced tended to move longer distances

than the others. Finally, while the point estimate for living in an urban area changes only a little, it now becomes statistically significant for men.

5.3 Conditioning on Post-War Outcomes

We next study how the coefficients for the displacement status change after controlling for post-war labor market fixed-effects. Columns (4) and (5) in Tables 3, 4, 5 and 6 present the estimates when we control for the full vector of indicator variables for the residence municipality in 1950 and 1970, respectively. In columns (6), we control for the 1970 residence municipality, sector of employment (54 categories) and socioeconomic status (5 categories). All specifications also control for the same pre-war observable characteristics as the results presented in column (3). As we discussed in Section 4, the coefficients in columns (4) to (6) do not have a causal interpretation, as they condition on post-war outcomes, which themselves are affected by the displacement. Rather, we consider them as descriptive statistics that are informative about the differences between a displaced and non-displaced person, who were similar before the war and were located in the same labor market after the war.

The main insight provided by these estimates is that displaced men had, on average, higher annual income than non-displaced men working in the same place, in the same industry and having the same socioeconomic status in 1970/71. However, this difference is considerably smaller than the estimate when we control only for pre-war observable characteristics. Furthermore, while the point estimates for pension income remain positive, they are no longer statistically significant once we condition on all these post-war characteristics. By contrast, displaced women had similar or slightly smaller income as non-displaced women regardless of the control variables added to the specification.

We return to the possible explanations for these rather surprising results in Section 7. Before turning to more speculative discussion, however, we collect the remaining facts and perform robustness checks. With regard to the remaining estimates, most are almost unaffected for controlling for the

post-war labor market fixed-effects. Perhaps the most interesting result is that displaced men were *less* likely to change their occupation between 1950 and 1970 than non-displaced men who were working in the same municipality and industry and having the same socioeconomic status in 1970. While the conditional results for industry/occupation mobility between 1950 and 1970 are weak in terms of statistical significance, these findings are consistent with the results above, suggesting that the impact of displacement on the pace of moving from traditional to modern occupations was predominantly a short-term phenomenon, although it led to persisting differences. Another interesting finding is that once we condition for the post-war residence municipality, displaced men are not more likely to be employed in the formal labor market than non-displaced men.

5.4 Heterogeneity of the Effect

We now turn to the question of whether the impact of forced migration varies across different groups of the displaced. We first look at the heterogeneity of the impact across age groups. Such heterogeneity could occur, for instance, because the younger cohorts have greater incentives to respond to the displacement by acquiring more education or learning a new profession. To some extent, the results support this hypothesis. Columns (1) to (4) in Table 7 present the estimates from specifications including an interaction term between age and displacement status as well as controls for pre-war observable characteristics. The results suggest that younger displaced were more likely to shift to higher income industries. That is, the estimates indicate that forced migration increased the 1970 imputed income by 0.15 log-points (standard error 0.03) among those displaced at age 14 in comparison to .03 log-points (standard error 0.04) among those displaced at age 32. Results for women are similar. Furthermore, the impact on being employed in the formal labor market is estimated to be 7.4 percentage points (standard error 2.3) among the youngest men, in comparison to -1.1 percentage points (standard error 2.9) among the oldest. However, the interaction terms are not statistically significant for actual income.

Another likely dimension along which the impact of displacement could vary is the pre-war sector of employment. The results above suggest that one important mechanism through which income effect occurred was a shift from traditional (rural) sectors to modern (urban) industries. The remaining columns in Table 7 support this hypothesis. We find that the interaction terms between displacement status and a dummy for working in agriculture before the war are significant for both imputed and actual income. They are also large in magnitude. For example, the point estimates for men suggest that the impact of forced migration on actual log income was 0.11 log-points larger for those working in agriculture in 1939 than for those working in other industries. For women, we find a positive effect on actual income *only* for those who worked in agriculture before the war. However, we do not find statistically significant evidence on that the impact on pension income would have differed between farmers and the rest of the population.

These results imply that the displacement had very different effects for persons differing in at least two observable pre-war characteristics. Another way to study effect heterogeneity is to look at the impact on different parts of the income distribution. To do this, Table 8 reports estimates from quantile regressions of income on displacement status. There is some variation in the results, depending on the dependent variable and whether we look at men or women. On balance, however, the results suggest that the impact is largest in the lower end of the income distribution and gradually fades away as we move towards higher quantiles. This is in particular the case for men and log income, where the lowest decile of income among the displaced is 0.44 log points higher than for the non-displaced, compared to a difference in medians of 0.23 log points and an insignificant difference of 0.04 points for the ninth decile. For women, the log income regressions yield no statistically significant differences, although the point estimates are decreasing across the distribution. For pension income, the lowest decile and quartile suggest higher incomes among the displaced. Furthermore, looking at the coefficient for level of taxable income, the median regression now suggests that forced migration had a positive albeit small income effect also among women.

6 Robustness Checks

6.1 Attrition

Our first robustness check considers the possibility that those observed in 1970 may not be a representative sample of those who were displaced during the war. It is conceivable, for instance, that proximity to the battles or stress caused by forced migration might have led to higher mortality rates among the displaced than the rest of population. On the other hand, the increased mobility caused by the displacement could have been reflected in increased propensity to emigrate, in particular given that a large share of the cohort we study left Finland during the 1950s and 1960s.¹⁰ Furthermore, one might argue that those with the weakest earnings potential may have been the most likely to die or emigrate. Note that if attrition rates were higher among the displaced and if those leaving the sample were negatively selected, our baseline estimates would be biased upwards.

We address this question in two ways. First, we study attrition rates by 1939 residence municipality. As discussed in more detail in the Appendix, we do this by comparing the number of individuals in the microdata to pre-war data on the population of municipalities. Table 9 reports the results from regressing number of individuals observed in 1970 in our microdata by their 1939 residence municipality on total pre-war population of this municipality, a dummy indicating that the municipality was ceded after the war and some pre-war municipality characteristics. The results suggest that there was no systematic differences in attrition rates between the displaced and non-displaced. We acknowledge, however, that the estimates are rather imprecise.

Our second approach is to ask how the results would change *if* important non-random attrition was present. Specifically, we consider the hypothetical case where (a) attrition among the displaced was five percentage points higher than among the non-displaced, and (b) this attrition was extremely negatively selected in a sense that those with the lowest earnings capacity

¹⁰ According to Statistics Finland, net outmigration between 1945 and 1970 was some 270,000 individuals. By far the most important destination was the neighboring Sweden.

are not present in 1970. To be clear, we do not claim that such attrition took place. Rather, the idea is to consider a scenario that would be particularly worrying for our key conclusions. The five percent attrition rate is arbitrary, but is intended to represent a high and therefore conservative rate of differential attrition.

Under the assumption that this type of attrition occurred, consistent estimates can be obtained by running the regressions on a sample omitting the lowest five percent of the non-displaced.¹¹ The second row of Table 10 presents these estimates for key outcome variables. As expected, the figures are smaller than the baseline estimates reproduced at the first row. Yet, we still find that displacement increased long-term income among men.

In short, we find no evidence of non-random attrition. Furthermore, we show that even if large differences in attrition rates were present and attrition was extremely negatively selected, the results would remain qualitatively intact. Thus we conclude that our main findings are not driven by non-random attrition.

6.2 Regional Price Variation

We next turn to the question whether the increase in nominal income among the displaced translates into an increase in consumption possibilities. This question is motivated by the notion that the displaced were more likely to move to urban areas. Thus, at least part of the gains in nominal income could have been lost due to higher prices, in particular for housing. As a consequence, part of the surplus created by migration would be transferred to landlords in the areas attracting many displaced.

However, this does not seem to be an important part of the story. The third row of Table 10 reports estimates for the impact on real income, which

¹¹See Angrist et al. (2006) for formal discussion. The key insight is that under the assumption of monotone treatment response – that becoming displaced never decreased the likelihood to die or to emigrate before 1970 – dropping the lower tail of the control distribution and running OLS for the remaining sample provides an estimate of lower bound of the treatment effect. Furthermore, under a relatively mild rank-preservation restriction this procedure point identifies the treatment effect. Similarly, the baseline estimates provide an upper bound.

is obtained by dividing the monetary outcomes measured in 1971 by the local price index.¹² Interestingly, we find that the impact on real income was even slightly larger than the impact on nominal income. This finding is a bit surprising, but partly reflects the fact that the most expensive rural municipalities are in Northern Finland, while most displaced were settled in the south.

6.3 Conditioning on Other Post-War Characteristics

We next study whether our main results are affected by conditioning on further observable characteristics. One unfortunate feature of the data is that they do not contain information about pre-war educational attainment. However, we have a good measure for level of education in 1950. Given that forced migration could affect the availability of schooling or the decision to acquire education, we chose not to control for 1950 education in our baseline specification. However, as the results reported in Table 10 show, conditioning on education has no effect on the estimates – cf. the first- and fourth-row estimates. These results strongly imply that differences in education are not driving our results.

Another possibility is studied in the fifth row of Table 10. The purpose of these estimates is to address a hypothesis that wealth effects give rise to the higher 1971 income among the displaced. Since the settlement policy provided only partial compensation for wealth lost, higher income among the displaced could result from the displaced working harder in order to recover from this shock rather than their higher productivity. Studying the relevance of this possibility is difficult because of the limited amount of information on wealth available in the data. Yet, some observed variables are reasonable proxies for wealth. Most importantly, we observe the amount of agricultural land owned by the household in 1950. Other available wealth proxies are the number of rooms per persons in the residence and dummies for living in

¹²These price data were collected in 1971 cost of living study that collected prices of 135 items from each municipality. A local price index was calculated based on consumption shares estimated in the 1966 Household Budget Survey. According to this study variation in local price level was mainly due to differences in (quality adjusted) rental prices. (Statistics Finland, 1972)

an owner-occupied dwelling and employing a servant, all measured in 1950. Conditioning on these variables has little impact on the estimates, suggesting that differences in post-war wealth, at least to the extent that we observe them, are not driving the results.

In addition to these alternative specifications, we have experimented with controlling for alternative industry/occupation definitions in the specifications leading to the estimates reported in columns (6) in Tables 3, 4, 5 and 6. The estimates reported in the tables use the 1939 industry coding, which is constructed from the richer categorization available only for 1970. However, it makes no difference whether we control for the baseline 54 industry categories or for the 274 categories available in the 1970 coding. Similarly conditioning on the 312 occupational categories available for 1970 (or for both the occupation and industry) has little impact on the estimates.

6.4 Measurement Issues

One possible objection to the results concerning the income effect is that taxable income may be a poor measure of true consumption possibilities or productivity. In particular, one could hypothesize that an important part of consumption among farmers would consist of consuming part of their own harvest and thus would not be recorded as part of taxable income. Since the displaced were more likely to move from agriculture to other sectors, such under-measurement of agricultural income would over-estimate the impact of the displacement. However, as reported in the fourth row of Table 10, running the regressions for a sample excluding those working in agriculture in 1970 yields estimates that are virtually identical to those for the whole sample. Thus, systematic measurement error, at least in this dimension, does not account for our results.

6.5 Alternative Control Areas

So far we have studied the impact of the displacement comparing the displaced to the entire population. In principle, one could argue that people living far away from the ceded area would systematically differ from those

living close to the ceded area and thus the former should not be included in the analysis. In the two last rows of Table 10 we checked whether the estimates are sensitive to the choice of the comparison region by including in the sample only those who lived in 1939 close to the post-war border. As the ceded areas extended 187 kilometers east from the post-war border we first limit the comparison region to municipalities that were no further than 187 kilometers west from the border. In the bottom row we limit both treatment and comparison groups to those living within 50 kilometers from the new border. The point estimates change very little, but the restrictions do make the estimates less precise.

In summary, the data strongly suggests that forced migration increased the income of displaced men. For women, the results are somewhat mixed. However, the estimates are remarkably stable across various specifications and alternative sample selection criteria. We next turn to discuss more informally the implications of these results.

7 Discussion

Our results show that forced migration increased the long-term economic outcomes of the displaced persons. Thus, at a minimum, the Finnish experience in settling 11% of the population illustrates that successful resettlement policy is possible. However, the finding that the displaced did significantly better than the rest of the population calls for explanations beyond resettlement policies. We next suggest some possibilities that are consistent with our empirical findings. These explanations are inevitably more speculative than the evidence presented in the previous sections.

The first part of our explanation is based on the idiosyncratic factors related to a specific historical period. The post-war years in Finland were an era of rapid industrialization. While half of the persons in our sample were employed in agriculture in 1939, this share had declined to a third among the non-displaced by 1970. The change was considerably faster among the displaced. Even though the displaced farmers were settled in farming communities and given an opportunity to continue farming, their farms were

small and often provided low standards of living. In addition, the displaced had fewer non-pecuniary reasons to stay in their assigned settlement areas. The resulting faster transformation from agrarian to modern occupations partially explains higher long-term income among the displaced.

However, we find that the displaced also earned more than those living in similar regions and employed in similar occupations after the war. This is substantially harder to explain. In our view, the most plausible explanation again has to do with lower mobility costs and the resulting higher mobility across regions and sectors. As shown above, the displaced were less likely to live in owner-occupied housing and were more likely to move not only from ceded areas to the rest of Finland, but also from the initial placement areas to other regions. A substantial fraction of these moves is likely to be a response to better employment opportunities. Higher mobility may have led to better matching between workers and jobs and hence to higher income among the displaced even within a labor market. As the displaced moved from the agrarian sector to industrial occupations before others, they also had a first-mover advantage and more time to find the best matches. Furthermore, earnings growth or return to tenure may have been higher in the modern sector. Since the displaced moved to modern occupations earlier, they simply may have enjoyed faster earnings growth for more years than similar non-displaced persons. Unfortunately, we do not have data on income in 1950, or indeed any of the intervening years, and thus we are unable to examine this hypothesis in detail.

The fact that women appear to have experienced much smaller gains from displacement warrants some discussion. If the economic advantages from early moves are driven by better labor matches, married women may not have been able to exploit those as effectively if the main bread-winner was the male. Thus, if women were more likely to be the secondary bread-winner in the family, moves were more likely driven by the better labor market matches of their husbands.

The final question concerns the general lessons this particular displacement provides for public policy. In our view, the results inform at least two policy debates. First, our findings provide some guidance on the question

on how displaced persons should be helped. The resettlement policy studied here was generous, providing both land and monetary compensation for lost property. The displaced farmers had the option of not taking up the offered land and were free to sell it – and many did. The displaced urban population received monetary compensation and was free to choose where to live. Whether the later economic success of the displaced should be viewed as a result of increased mobility at a time of rapid structural change or as the result of a successful resettlement policy is not evident. Arguably, it is a bit of both. Yet, in the absence of a settlement plan and compensation for lost property, the displaced might have stayed long in the evacuation areas and formed an impoverished underclass. The settlement plan did not lock them into the traditional sector, but provided the means to start over.

Our results are also relevant for regional policy. We find that moving early into the modern sector – which often entailed geographic mobility – was financially rewarding for the displaced. The hypothesized explanation, their lower pecuniary and non-pecuniary moving costs, suggests that policies aimed at lowering the costs of migration may be an effective way to reduce regional disparities. Thus, policies encouraging mobility, such as subsidizing voluntary moves or discontinuing tax subsidization of owner-occupied housing, might be more efficient than place-based policies that attempt to help people who stay in the economically declining regions.

8 Conclusions

Post World War II evacuation of the Eastern parts of Finland created an situation where 430,000 persons had to relocate to the remaining parts of the country. We have exploited this historical episode to study the impact of displacement on those who were forced to migrate. Our findings indicate that being displaced had significant positive effects on long-term economic outcomes.

Since we focus on migrants who had no choice but to move, the setup allows us to avoid the fundamental selectivity problems present in much of the previous research on the economic effects of migration. Thus, we provide

a consistent estimate of returns to (forced) migration on an average person in the data. The results suggest that higher mobility enhanced the efficiency of resource allocation in post-war Finland.

Much migration consists of either voluntary moves or population displacements in very disintegrated countries. Yet, there are many important situations resembling the one studied here. Examples of such displacements include large-scale public infrastructure projects and conflict-induced displacements in many parts of the world, as well as displacements caused by natural disasters permanently turning some areas uninhabitable. These situations call for active settlement policies, possibly financed by rich countries. The Finnish experience of resettling the displaced and their subsequent economic success may hold useful lessons for dealing with these forced migrations.

References

- AGRAWALA, S., T. OTA, A. A. UDDIN, J. SMITH, AND M. V. AALST (2003): *Development and Climate Change in Bangladesh: Focus on Coastal Flooding and the Sundarbans*, OECD, Environment Directorate and Development Co-operation Directorate.
- ANGRIST, J., E. BETTINGER, AND M. KREMER (2006): “Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia,” *American Economic Review*, 96, 847–862.
- ANGRIST, J. D. AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist’s Companion*, Princeton University Press.
- BORJAS, G. (2001): “Does Immigration Grease the Wheels of the Labor Market?” *Brookings Papers on Economic Activity 2001*, 69–119.
- CZAIKA, M. AND K. KIS-KATOS (2007): “Civil Conflict and Displacement. Village-Level Determinants of Forced Migration in Aceh,” HiCN Working Paper 32.
- GOTTSCHANG, T. (1987): “Economic Change, Disasters, and Migration: The Historical Case of Manchuria,” *Economic Development and Cultural Change*, 35, 461.
- GREENWOOD, M. (1997): “Internal Migration in Developed Countries,” in *Handbook of Population and Family Economics*, ed. by M. Rosenzweig and O. Stark, Amsterdam: Elsevier, vol. 1, 647–720.
- IBÁÑEZ, A. M. AND C. E. VÉLEZ (2008): “Civil Conflict and Forced Migration: The Micro Determinants and Welfare Losses of Displacement in Colombia,” *World Development*, 36, 659–676.
- KONDYLIS, F. (2007): “Conflict Displacement and Labour Market Outcomes: Evidence from post-war Bosnia and Herzegovina,” CEP Discussion Paper 778.

- (2008): “Agricultural Households and Conflict Displacement: Evidence from a Policy Intervention in Rwanda,” *Forthcoming in Economic Development and Cultural Change*.
- LUCAS, R. (1997): “Internal Migration in Developing Countries,” in *Handbook of Population and Family Economics*, ed. by M. Rosenzweig and O. Stark, Amsterdam: Elsevier, vol. 1, 721–798.
- PIHKALA, K. U. (1952): “The Land Settlement Program of Finland,” *Land Economics*, 28, 147–159.
- SCHULTZ, T. (1971): “Rural-Urban Migration in Colombia,” *Review of Economics and Statistics*, 53, 157–163.
- SJAASTAD, L. (1962): “The Costs and Returns of Human Migration,” *The Journal of Political Economy*, 70, 80–93.
- STATISTICS FINLAND (1942): *IV Varallisuustilastoa: Tulo ja omaisuusverotilastoa vuodelta 1938*, Helsinki: Valtioneuvoston kirjapaino.
- (1953): *IV Varallisuustilastoa: Tulo ja omaisuusverotilastoa vuodelta 1950*, Helsinki: Valtioneuvoston kirjapaino.
- (1972): *Vuoden 1971 kuntien kalleustutkimus*, Studies 19, Statistics Finland.
- (1996): *The Finnish Longitudinal Census Data File*, Helsinki: Statistics Finland.
- TARKKA, J. (1988): “Paasikiven aika,” in *Suomen Historia 8*, ed. by J. Tarkka, Espoo: Weilin ja Göös, 1–2.
- UNHCR (2008): *2007 Global Trends: Refugees, Asylum-seekers, Returnees, Internally Displaced and Stateless Persons*, UNHCR.
- WARIS, H., V. JYRKILÄ, K. RAITASUO, AND J. SIIPPI (1952): *Siirtoväen sopeutuminen*, Otava.

WORLD COMMISSION ON DAMS (2000): *Dams and Development: A New Framework for Decision-Making*, Earthscan Publications.

A Appendix

A.1 Constructed Variables

Imputed income

We construct imputed income variable for each individual using the coefficient estimates from a regression of log annual taxable income in 1971 on a full set of age, socioeconomic status and industry dummies for a sample aged 18–59 in 1971. In essence, this procedure assigns each industry–age–socioeconomic status combination in all years a value representing the expected income of persons of the same socioeconomic status working in that particular industry in 1970/71. The regressions are run separately for men and women.

While providing a useful summary measure, this procedure has several shortcomings. Most importantly, imputed income is not informative on whether there was intra-industry income differences between ceded areas and the rest of the country. Unfortunately, there is little we can do about this problem. Second, the wage structure may have changed substantially between 1939 and 1971. We address this possibility by using tables listing mean taxable income in 38 industry-occupation-socioeconomic status groups for 1950 (Statistics Finland, 1953) and 12 industry-living in urban area groups for 1939 (Statistics Finland, 1942) and assign each individual the mean income of their reference group.

Lives in urban area

Statistics Finland 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, Viipuri (ceded), Tampere, Turku, Vaasa,

Lahti, Oulu, Kuopio, Kotka, Kemi, Pori, Lappeenranta, Mikkeli, Rauma, Hämeenlinna, Jyväskylä, Kokkola, Savonlinna, Hanko, Porvoo, Kajaani, Pietarsaari, Joensuu, Hamina, Sortavala (ceded), Käkisalmi (ceded), Loviisa, Tammisaari, Iisalmi, Raahe, Uusikaupunki, Heinola, Kristiinankaupunki, Tornio, Kaskinen, Uusikaarlepyy and Naantali.

Taxable income per capita in the residence municipality

For 1939 and 1950, this variable is constructed by dividing the sum total of taxable income (*veroäyri*) by the number of residents in the municipality. For 1970, we calculate average income from the microdata using 1971 data on taxable income.

A.2 Attrition

The aim is to learn how the available pre-war population data relate to the number of individuals observed in our sample. This relationship is by definition

$$N_{j70} = o_j S_j P_{j39} \xi_j \quad (5)$$

where N_{j70} is the number of individuals observed in microdata in 1970 who lived in municipality j in 1939, o_j is the sampling rate for this municipality¹³, S_j is the survival rate, P_{j39} is population in 1939 and ξ_j captures measurement error in P_{j39} . The measurement error results from our restricted interest to particular age groups, while the municipality data are for the entire population. We model the survival rate as

$$\ln S_j = \ln \bar{S}_j + \beta D_j + X_j \gamma + v_j \quad (6)$$

where $\ln \bar{S}_j$ is the logarithm of the mean survival rate among municipalities that remained part of Finland, D_j is a dummy indicating that the municipality was ceded after the war, X_j is vector of pre-war characteristics such as

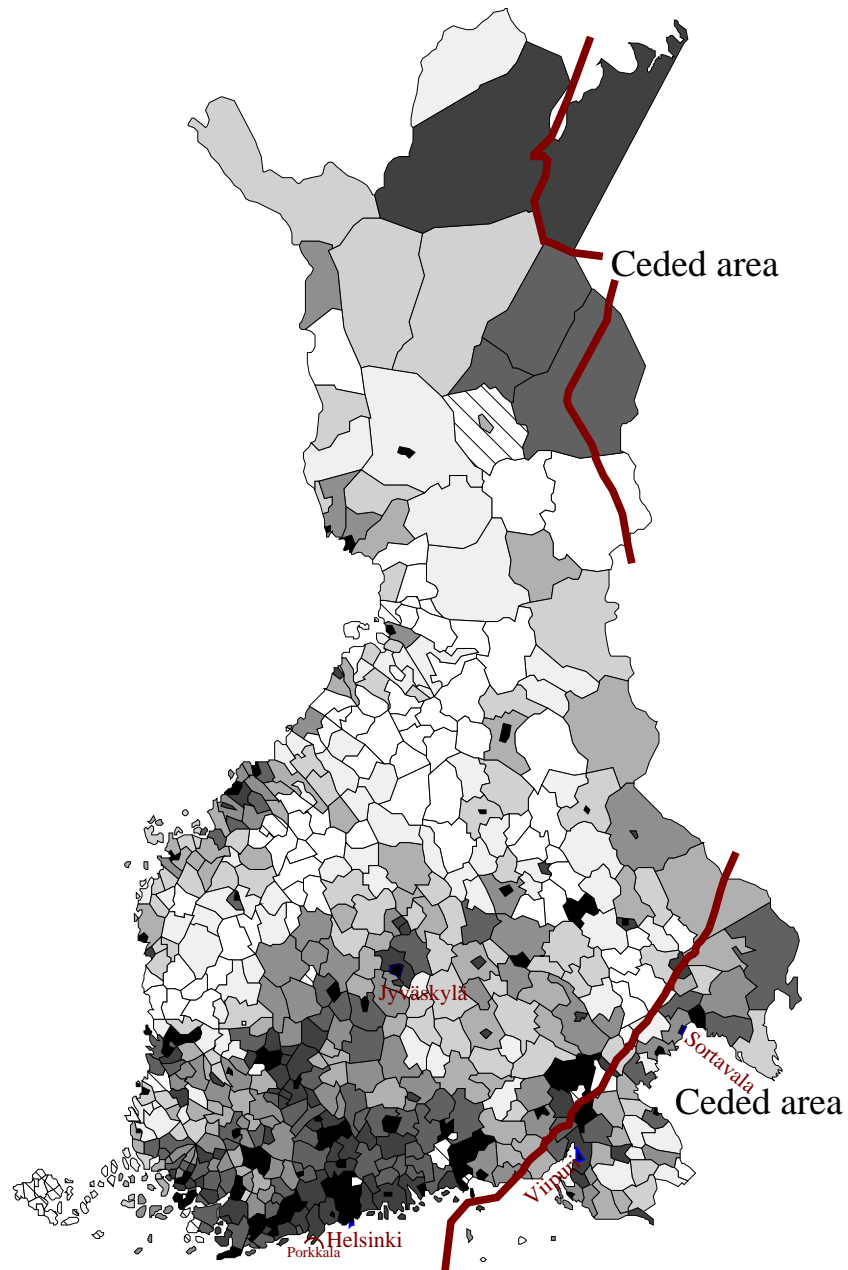
¹³As the sampling was made by taking every tenth file, this will vary across municipalities

the age structure and u_j is an error term. Taking logs of (5) and substituting with (6), we get an estimation equation

$$\ln N_{j70} = \alpha + \beta D_j + \theta \ln P_{j39} + X_j \gamma + \epsilon_{j70} \quad (7)$$

where $\alpha = \ln \bar{o} + \ln \bar{S}_j$, and $\epsilon_{j70} = (\ln \bar{o} - \ln o_j) + v_j + \ln \xi_j$. Under the assumption that ϵ is uncorrelated with the covariates, parameter β is informative on whether attrition among the displaced differs from the rest of the population. Furthermore, the fact that (5) is an accounting relation also suggests a natural specification check: we should find that $\theta = 1$.

Figure 1: Ceded Areas and Taxable Income per capita, 1938



Note: Darker shades indicate higher taxable income per capita.

Table 1: Sample means

	1939		1950		1970	
	Disp.	Other	Disp.	Other	Disp.	Other
Age	22.9	22.8	33.9	33.8	53.8	53.7
Female	0.53	0.53	0.53	0.53	0.55	0.54
Lives in urban area	0.24	0.27	0.33	0.30	0.48	0.41
Native tongue Swedish	0.01	0.09	0.01	0.09	0.00	0.09
Public sector	na.	na.	0.20	0.14	0.15	0.13
Single	na.	na.	0.13	0.12	0.20	0.18
Owner occupied housing	na.	na.	0.43	0.54	0.71	0.76
Migrated	na.	na.	1.00	0.29	0.35	0.24
Taxable income per capita in the mun. of residence	46.8	50.3	688	635	16.0	14.7
Imp. log income (def. 1)	1.91	1.94	2.35	2.28	2.32	2.26
Imp. log income (def. 2)	2.93	2.92	5.40	5.34	na.	na.
Annual taxable income	na.	na.	na.	na.	14.5	13.2
<i>Socioeconomic status</i>						
entrepreneur	0.09	0.08	0.14	0.18	0.17	0.22
white collar worker	0.10	0.10	0.21	0.19	0.20	0.18
blue collar worker	0.27	0.34	0.47	0.44	0.37	0.33
assisting family member	0.16	0.14	0.16	0.17	0.04	0.06
unemployed/out of LF	0.38	0.35	0.02	0.02	0.22	0.21
Employed	0.47	0.52	0.82	0.81	0.74	0.73
<i>Sector (conditional on not unknown/missing)</i>						
Agriculture	0.45	0.47	0.28	0.38	0.22	0.32
Manufacturing etc.	0.19	0.21	0.26	0.26	0.26	0.23
Construction	0.07	0.06	0.11	0.09	0.12	0.10
Service etc.	0.30	0.27	0.34	0.27	0.40	0.36
Sector unknown	0.40	0.37	0.05	0.05	0.23	0.22
<i>Education</i>						
Primary	na.	na.	0.78	0.75	na.	na.
Secondary (9 years)	na.	na.	0.07	0.06	na.	na.
Secondary (12 years)	na.	na.	0.03	0.04	na.	na.
None / unknown	na.	na.	0.12	0.15	na.	na.
Observations	2,558	20,213	2,558	20,213	2,326	18,454

Note: "Migrated" refers to a change in residence municipality between 1939 and 1950 (fourth and fifth columns), and between 1950 and 1970 (sixth and seventh columns). Taxable income per capita in the municipality of residence are in nominal terms and not comparable across time periods. Imputed earnings (definition 1) assigns the 1971 earnings distribution to all years on industry-socioeconomic status-living in an urban area level. Imputed earnings (definition 2) is based on 1939 and 1950 actual earnings distributions in nominal terms and not comparable across time periods. See the Appendix for details.

Table 2: Pre-War Differences

Dependent Variable	Mean	Displacement status		Obs.
		w/o Cov. (2)	w/ Cov. (3)	
<i>A: Men</i>				
Employed in the Formal Labor Market	0.71	-0.033 (0.034)	-0.038 (0.023)	10,673
Imputed log Earnings (def. 1)	2.48	0.001 (0.098)	0.008 (0.031)	10,270
Imputed log Earnings (def. 2)	2.87	0.015 (0.032)	0.027* (0.017)	8,590
log Taxable income per capita in the municipality of residence	3.54	0.089 (0.222)	-0.076 (0.122)	10,673
<i>B: Women</i>				
Employed in the Formal Labor Market	0.33	-0.063 (0.052)	-0.042** (0.018)	12,098
Imputed log Earnings (def. 1)	1.47	-0.053 (0.102)	-0.031 (0.024)	11,914
Imputed log Earnings (def. 2)	2.98	0.010 (0.059)	0.003 (0.018)	5,603
log Taxable income per capita in the municipality of residence	3.63	0.022 (0.239)	-0.114 (0.128)	12,098

Note: OLS estimates of future displacement status and pre-war outcomes. Each coefficient comes from a separate regression. Standard errors (in parentheses) robust to clustering at municipality of residence in 1939 level. Specification for column (3) controls for age, latitude and longitude of the 1939 residence municipality, and indicator variables for speaking Swedish as ones mother tongue and living in urban area in 1939. ***, **, * indicate statistical significance at 1%, 5%, 10% level, respectively.

Table 3: Short-Term Outcomes (Men)

Dependent Variable	Mean	The Effect of Displacement		Conditioning on Post-War Labor Market and Covs			Obs.
		w/o Covs (2)	w/ Covs (3)	1950a (4)	1970a (5)	1970b (6)	
Emp. in Formal Labor Market	0.90	-0.02 (0.02)	-0.02 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	10,673
Imp. log Income (1)	2.7	0.06 (0.08)	0.07*** (0.02)	0.02 (0.02)	0.06** (0.03)	0.04* (0.02)	10,509
Imputed log Income (2)	5.3	0.09** (0.04)	0.11*** (0.02)	0.07*** (0.02)	0.10*** (0.03)	0.09*** (0.03)	9,543
Lives in urban area	0.29	0.04 (0.10)	0.06 (0.04)	.	0.02 (0.04)	0.02 (0.04)	10,673
Changed Sect. (1-dig)	0.42	0.10*** (0.01)	0.10*** (0.02)	0.04 (0.02)	0.09*** (0.03)	0.09*** (0.03)	10,673
Changed Sect. (2-dig)	0.48	0.09*** (0.02)	0.11*** (0.02)	0.06*** (0.02)	0.11*** (0.03)	0.11*** (0.03)	10,673

Note: Impact of being displaced. OLS estimates on 1950 outcomes, standard errors (in parentheses) robust to clustering at 1939 residence municipality level. Each coefficient comes from a separate regression. Covariates: age, imputed earnings in 1939, latitude, longitude and log taxable income per capita of the 1939 residence municipality, and indicator variables for speaking Swedish as ones mother tongue and living in city or market town in 1939. ***, **, * indicate statistical significance at 1%, 5%, 10% level, respectively.

Table 4: Short-Term Outcomes (Women)

Dependent Variable	Mean	The Effect of Displacement		Conditioning on Post-War Labor Market and Covs			Obs.
		w/o Covs (2)	w/ Covs (3)	1950a (4)	1970a (5)	1970b (6)	
Emp. in Formal Labor Market	0.73	0.04 (0.05)	0.04* (0.02)	-0.00 (0.02)	0.02 (0.02)	0.02 (0.02)	12,098
Imp. log Income (1)	1.9	0.08 (0.10)	0.11*** (0.03)	0.04 (0.03)	0.06** (0.03)	0.06** (0.02)	11,982
Imputed log Income (2)	5.4	0.03 (0.03)	0.07*** (0.02)	0.04* (0.02)	0.04* (0.02)	0.04* (0.02)	8,765
Lives in urban area	0.32	0.01 (0.10)	0.06 (0.04)	.	0.01 (0.03)	0.01 (0.03)	12,098
Changed Sect. (1-dig)	0.69	0.05*** (0.01)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	12,098
Changed Sec. (2-dig)	0.73	0.04** (0.02)	0.02 (0.02)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)	12,098

Note: Impact of being displaced on 1950 outcomes. OLS estimates, standard errors (in parentheses) robust to clustering at 1939 residence municipality level. Each coefficient comes from a separate regression. Covariates: age, imputed earnings in 1939, latitude, longitude and log taxable income per capita of the 1939 residence municipality, and indicator variables for speaking Swedish as ones mother tongue and living in city or market town in 1939. ***, **, * indicate statistical significance at 1%, 5%, 10% level, respectively.

Table 5: Long-Term Outcomes (Men)

Dependent Variable	Mean	The Effect of Displacement		Conditioning on Post-War Labor Market and Covs			Obs.
		w/o Covs (2)	w/ Covs (3)	1950 (4)	1970a (5)	1970b (6)	
Emp. in Formal Labor Market	0.82	-0.00 (0.01)	0.03** (0.02)	0.02 (0.02)	0.01 (0.02)	.	9,496
Imputed log Income (def. 1)	2.7	0.06 (0.05)	0.10*** (0.02)	0.04 (0.02)	0.03 (0.02)	.	9,494
Taxable Income ('000 marks)	20.8	2.26 (2.17)	2.64*** (0.99)	2.03** (0.96)	2.30** (0.96)	1.46** (0.72)	9,496
log Income	2.8	0.17** (0.08)	0.24*** (0.04)	0.12*** (0.04)	0.11** (0.04)	0.08** (0.03)	8,572
log Pensions (various years)	2.4	0.13* (0.07)	0.13*** (0.05)	0.12** (0.05)	0.11** (0.05)	0.07 (0.04)	7,383
Changed Sec. (1-dig)	0.46	0.04** (0.02)	-0.00 (0.03)	0.01 (0.03)	-0.01 (0.03)	-0.01 (0.02)	9,496
Changed Sec. (2-dig)	0.56	0.04** (0.02)	-0.01 (0.03)	0.02 (0.03)	0.01 (0.03)	0.02 (0.02)	9,496
Changed Occ. (1-dig)	0.44	0.03* (0.02)	-0.02 (0.02)	-0.00 (0.02)	-0.04 (0.03)	-0.04* (0.02)	9,496
Lives in urban area	0.40	0.08 (0.07)	0.08* (0.05)	0.04* (0.02)	.	.	9,496
Migrates 1950–1970	0.36	0.11*** (0.02)	0.04 (0.03)	0.02 (0.03)	0.02 (0.04)	0.01 (0.04)	10,510
Distance Migrated	130.0	10.97 (11.66)	61.42* (36.34)	0.98 (11.88)	38.34 (27.75)	35.86 (28.26)	2,535

Note: Impact of being displaced. OLS estimates, standard errors (in parentheses) robust to clustering at 1939 residence municipality level. Income measured in 1971 except pensions, which are measured as the first observed pension income in 1975, 1980, 1985 or 1990. Other outcomes measured in 1970. Each coefficient comes from a separate regression. Covariates: age, imputed earnings in 1939, latitude, longitude and log taxable income per capita of the 1939 residence municipality, and indicator variables for speaking Swedish as ones mother tongue and living in city or market town in 1939. Estimates reported in column (4) are also conditional on 1950 residence municipality, estimates presented in column (5) are conditional on 1970 residence municipality, estimates presented in column (6) are conditional on 1970 residence municipality, industry and socioeconomic status fixed-effects. ***, **, * indicate statistical significance at 1%, 5%, 10% level, respectively.

Table 6: Long-Term Outcomes (Women)

Dependent Variable	Mean	The Effect of Displacement		Conditioning on Post-War Labor Market and Covs			Obs.
		w/o Covs (2)	w/ Covs (3)	1950a (4)	1970a (5)	1970b (6)	
Emp. in Formal Labor Market	0.66	0.01 (0.02)	0.06*** (0.02)	0.03* (0.02)	0.03* (0.02)	.	11,284
Imputed log Income (def. 1)	1.9	0.06 (0.06)	0.11*** (0.03)	0.06*** (0.02)	0.04** (0.02)	.	11,280
Taxable Income ('000 marks)	7.2	0.48 (1.02)	-0.02 (0.48)	0.09 (0.57)	-0.43 (0.52)	-0.80* (0.45)	11,284
log Income	2.3	-0.01 (0.08)	-0.01 (0.05)	0.08 (0.05)	-0.01 (0.05)	-0.05 (0.04)	5,664
log Pensions (various years)	1.8	0.09 (0.07)	0.11** (0.05)	0.09** (0.04)	0.08** (0.04)	0.06 (0.04)	8,554
Changed Sec. (1-dig)	0.54	0.05*** (0.01)	0.02 (0.02)	0.02 (0.02)	-0.01 (0.02)	0.01 (0.02)	11,284
Changed Sec. (2-dig)	0.64	0.05** (0.02)	0.02 (0.02)	0.01 (0.02)	-0.01 (0.02)	0.00 (0.02)	11,284
Changed Occ. (1-dig)	0.51	0.03* (0.01)	0.03* (0.02)	0.05** (0.02)	0.02 (0.02)	0.01 (0.02)	11,284
Lives in urban area	0.43	0.07 (0.07)	0.07 (0.05)	0.05** (0.02)	.	.	11,284
Migrates 1950–1970	0.32	0.10*** (0.02)	0.05* (0.03)	0.04 (0.03)	0.02 (0.04)	0.02 (0.04)	11,916
Distance Migrated	129.3	11.83 (8.85)	56.91* (30.45)	2.78 (11.23)	29.61 (19.96)	28.79 (19.95)	2,989

Note: Impact of being displaced. OLS estimates, standard errors (in parentheses) robust to clustering at 1939 residence municipality level. Income measured in 1971 except pensions, which are measured as the first observed pension income in 1975, 1980, 1985 or 1990. Other outcomes measured in 1970. Each coefficient comes from a separate regression. Covariates: age, imputed earnings in 1939, latitude, longitude and log taxable income per capita of the 1939 residence municipality, and indicator variables for speaking Swedish as ones mother tongue and living in city or market town in 1939. Estimates reported in column (4) are also conditional on 1950 residence municipality, estimates presented in column (5) are conditional on 1970 residence municipality, estimates presented in column (6) are conditional on 1970 residence municipality, industry and socioeconomic status fixed-effects. ***, **, * indicate statistical significance at 1%, 5%, 10% level, respectively.

Table 7: Impact of Displacement, Interacted with Age and Working in Agriculture in 1939

Dependent Variable	Works in Agriculture in 1939											
	Age (deviation from mean, years)				Men				Women			
	Men		Women		Men		Women		Men		Women	
	Disp. (1)	Interaction (2)	Disp. (3)	Interaction (4)	Disp. (5)	Interaction (6)	Disp. (7)	Interaction (8)	Disp. (9)	Interaction (10)	Disp. (11)	Interaction (12)
Employed (1950)	-0.017 (0.014)	0.001 (0.003)	0.046* (0.024)	-0.004 (0.003)	0.002 (0.015)	-0.023 (0.024)	0.049** (0.023)	-0.000 (0.034)				
Employed (1970)	0.033** (0.016)	-0.005** (0.002)	0.058*** (0.016)	-0.003 (0.002)	0.032* (0.018)	0.011 (0.024)	0.052*** (0.017)	0.039 (0.030)				
Imputed log Income (1950)	0.068*** (0.022)	-0.001 (0.003)	0.111*** (0.034)	-0.007* (0.004)	0.037* (0.020)	0.072** (0.031)	0.110*** (0.033)	0.040 (0.042)				
Imputed log Income (1970)	0.093*** (0.022)	-0.007** (0.003)	0.109*** (0.026)	-0.006** (0.003)	0.053** (0.024)	0.097*** (0.033)	0.095*** (0.028)	0.093** (0.041)				
Income (1970)	2.602*** (0.980)	-0.111 (0.110)	-0.007 (0.482)	-0.043 (0.041)	1.922 (1.171)	2.002 (1.353)	-0.181 (0.524)	1.397** (0.628)				
log Income (1970)	0.239*** (0.044)	0.002 (0.005)	-0.010 (0.046)	-0.005 (0.007)	0.186*** (0.048)	0.114** (0.055)	-0.023 (0.046)	0.277*** (0.102)				
log Pensions (various years)	0.137*** (0.048)	0.005 (0.008)	0.115** (0.047)	0.005 (0.006)	0.090* (0.052)	0.069 (0.058)	0.112** (0.045)	0.018 (0.091)				

Note: Impact of being displaced and interactions with age (columns 1 to 4) and a dummy for working in agriculture in 1939. OLS estimates, standard errors (in parentheses) robust to clustering at 1939 residence municipality level. Controlling for age, imputed earnings in 1939, latitude, longitude and log taxable income per capita of the 1939 residence municipality, and indicator variables for speaking Swedish as ones mother tongue and living in city or market town in 1939. ***, **, * indicate statistical significance at 1%, 5%, 10% level, respectively.

Table 8: Quantile Regressions

	Quantile				
	.1	.25	.5	.75	.9
<i>A: Men</i>					
Taxable Income, 1971 (‘000 marks)	0.77 (0.63)	3.15*** (0.83)	3.90*** (0.90)	4.08*** (1.05)	0.65 (2.18)
Quantiles among non-displaced	1	6	17	27	41
log Income, 1971	0.44*** (0.12)	0.29*** (0.07)	0.23*** (0.05)	0.13*** (0.04)	0.04 (0.06)
log Pensions, several years	0.27*** (0.10)	0.08 (0.06)	0.11** (0.05)	0.13** (0.07)	0.08 (0.07)
<i>B: Women</i>					
Taxable Income, 1971 (‘000 marks)	.	.	0.41** (0.21)	0.81 (0.71)	1.01 (0.83)
Quantiles among non-displaced	0	0	0	13	21
log Income, 1971	0.12 (0.13)	0.05 (0.08)	-0.07 (0.06)	-0.05 (0.05)	-0.07 (0.06)
log Pensions, several years	0.27*** (0.10)	0.17** (0.07)	0.07 (0.05)	0.04 (0.05)	-0.05 (0.06)

Note: Impact of being displaced. Quantile regressions, bootstrapped standard errors (in parentheses) based on 1,000 replications and robust to clustering at 1939 residence municipality level. Pensions measured as the first observed pension income in 1975, 1980, 1985 or 1990. Controlling for age, imputed earnings in 1939, latitude, longitude, log taxable income per capita of the 1939 residence municipality and indicator variables for speaking Swedish as ones mother tongue and living in city or market town in 1939. ***, **, * indicate statistical significance at 1%, 5%, 10% level, respectively.

Table 9: Attrition

	$y = \log \# \text{observations, 1970}$			
	(1)	(2)	(3)	(4)
Ceded municipality	0.026 (0.114)	-0.002 (0.109)	-0.011 (0.109)	0.039 (0.139)
log Population, 1939	1.090 (0.045)	1.032 (0.044)	1.027 (0.044)	1.033 (0.046)

Note: Differences in attrition rates. The coefficients correspond to β (first row) and θ (second row) of equation (7), see the Appendix. Specifications: (1) no additional covariates, (2) controlling for log mean age, (3) controlling for age and share of Swedish-speaking population, (4) controlling for age, Swedish speaking population, longitude and latitude. All control variables are measured in the level of 1939 residence municipality.

Table 10: Robustness Checks

Specification	Imputed Income		Income		log Income	
	Men (1)	Women (2)	Men (3)	Women (4)	Men (5)	Women (6)
Baseline	0.10*** (0.02)	0.11*** (0.03)	2.64*** (0.99)	-0.02 (0.48)	0.24*** (0.04)	-0.01 (0.05)
Assuming selected attrition	0.06*** (0.02)	0.06*** (0.02)	1.75* (1.00)	-0.34 (0.50)	0.20*** (0.04)	-0.13*** (0.05)
Inflating with local prices	0.13*** (0.03)	0.14*** (0.03)	3.08*** (0.92)	0.22 (0.46)	0.26*** (0.04)	0.03 (0.04)
Conditional on 1950 Education	0.10*** (0.02)	0.11*** (0.03)	3.56*** (0.73)	0.16 (0.47)	0.25*** (0.04)	0.00 (0.05)
Conditional on 1950 Wealth	0.09*** (0.02)	0.09*** (0.02)	2.51*** (0.92)	-0.04 (0.44)	0.22*** (0.04)	-0.01 (0.04)
Excluding Agri- culture 1970	0.10*** (0.03)	0.07*** (0.02)	1.99* (1.17)	-0.67 (0.58)	0.22*** (0.05)	-0.04 (0.05)
Alternative area (187km)	0.07** (0.03)	0.07* (0.04)	3.55*** (1.14)	-0.03 (0.49)	0.21*** (0.06)	-0.03 (0.07)
Alternative area (50km)	0.09** (0.04)	0.06* (0.04)	3.68** (1.44)	-0.03 (0.54)	0.20*** (0.07)	0.09 (0.08)

Note: Impact of being displaced. OLS estimates, standard errors (in parentheses) robust to clustering at 1939 residence municipality level. Each coefficient comes from a different regressions. Each specification controls for age, imputed earnings in 1939, latitude, longitude and log taxable income per capita of the 1939 residence municipality, and indicator variables for speaking Swedish as ones mother tongue and living in city or market town in 1939. Additional controls: four categories of 1950 education (fourth specification); 17 dummies for the amount of agricultural land owned, rooms per resident and dummies for living in owner occupied premises and having a servant (fifth specification); ***, **, * indicate statistical significance at 1%, 5%, 10% level, respectively.