Agglomeration in the Periphery*

Matti Sarvimäki

June 2011

Abstract

This paper argues that agglomeration economies are important even in the rural periphery. The analysis focuses on the forced relocation of more than a tenth of the Finnish population after World War II. I use the details of the resettlement policy to construct instrumental variables for the wartime population growth rate in rural municipalities. The results suggest that an exogenous increase in local labor force had a positive effect on later population growth, industrialization and wages. These findings are consistent with the presence of agglomeration economies and inconsistent with other popular explanations for the spatial distribution of economic activity.

JEL: R12, J61, N94

Keywords: agglomeration economies, natural advantages, migration, spatial equilibrium

*This is a slightly revised version of SERC DP 80. I thank David Card, Steve Gibbons, Benjamin Faber, Kristiina Huttunen, William Kerr, Guy Michaels, Henry Overman, Steve Redding, Daniel Sturm, Marko Terviö, Roope Uusitalo, Olof Åslund and various seminar participants for helpful discussions and comments. I am particularly indebted to Aki Kangasharju who acted as a co-author in an early stage of this project. Tiina Hytönen, Nina Intonen and Miikka Rokkanen provided able research assistance. Ossi Kotavaara, Harri Antikainen and Jarmo Rusanen kindly constructed the railroad indicator. Financial support from the Yrjö Jahnsson Foundation is gratefully acknowledged. The Web Appendix is available at my website, currently www.hse-econ.fi/sarvimaki/. *Affiliations*: Aalto University, VATT, HECER and SERC/LSE. *Contact:* matti.sarvimaki@aalto.fi.

1 Introduction

Ever since Marshall (1890), economists have examined whether firms and workers become more productive when they are close to other firms and workers. Such externalities are now widely seen as the main reason for why cities exist. In this paper, I will argue that agglomeration economies are important also in the rural periphery.

I start by setting up a simple model that captures the three leading explanations for the spatial distribution of economic activity: the endowments of immobile factors, random growth processes and agglomeration economies. The model illustrates how migration shocks can be used to test for the presence of agglomeration economies. The intuition is the following. Consider a hypothetical experiment that settles a large number of workers into randomly chosen locations. Suppose that after the experiment, workers start migrating from the 'control' to the 'treatment' areas. This finding would be consistent with the agglomeration economies explanation and inconsistent with the immobile factors and random growth explanations.

Settling a large number of people to randomly chosen locations is, of course, ethically and practically infeasible. Thus I exploit a natural experiment that closely resembles the hypothetical experiment. After World War II, Finland ceded its eastern parts to the Soviet Union and relocated 11 percent of its population to the remaining parts of the country. The number of displaced farmers that each rural location received was determined by the amount of government owned land and the size distribution of private farms. A further source of variation was created by the decision of settling virtually no-one into the Swedish-speaking parts of the country. I use these features of the resettlement policy to construct instrumental variables for municipality-level population growth rate between 1939 and 1949. Under the assumption that the instruments are valid, this allows me to examine the causal effect of a labor supply shock on later outcomes.

The plausibility of my empirical strategy is supported by the fact that the identifying variation dates back to the Middle Ages. I will argue that while part of this variation persisted, its economic rationale had vanished by the early 20th century due to the expansion of population, the end of the Little Ice Age, revolution in transportation technology and the shift of the economic center from Stockholm to St Petersburg. Furthermore, I show that the instruments do not explain pre-war population growth rates, that the key results remain stable when I use each instrument individually and that the exclusion restriction would have to be violated by a large magnitude in order to alter the conclusions.

In line with the agglomeration economies explanation, I find that the resettlement shocks increased later population growth. According to the point estimates, a 10 percent increase in municipality's population due to the resettlement policy caused an additional 17 percent growth during the next five decades. This growth occurred due to increased internal migration from other rural areas. Furthermore, the resettlement shock led to an expansion of the non-primary sector and improved wages.

These results contribute to three branches of research. First, I add to the growing empirical literature on agglomeration economies. Previous work studying war-related shocks include Davis and Weinstein (2002, 2008), Brakman, Garretsen, and Schramm (2004), Bosker et al. (2007, 2008) and Redding, Sturm, and Wolf (forthcoming).[1] All of these studies examine negative shocks on cities. In contrast, I estimate the impact of a positive migration

[1]See also Acemoglu, Hassan, and Robinson (forthcoming) and Miguel and Roland (forthcoming) on the long-term effects of war-related shocks, and Hornbeck (2009) for the economic adjustment to an evironmental catastrophe. The broader empirical literature on agglomeration economies includes, but is not limited to, Ciccone and Hall (1996), Ellison and Glaeser (1997, 1999), Rosenthal and Strange (2001, 2003, 2008), Rappaport and Sachs (2003), Head and Mayer (2004a), Hanson and Xiang (2004), Duranton and Overman (2005), Hanson (2005), Amiti and Cameron (2007), Arzaghi and Henderson (2008), Redding and Sturm (2008), Combes et al. (2009), Partridge et al. (2009), Ellison, Glaeser, and Kerr (2010), Glaeser and Resseger (2010) and Greenstone, Hornbeck, and Moretti (2010). See Overman, Redding, and Venables (2003), Head and Mayer (2004b), Rosenthal and Strange (2004), Glaeser and Gottlieb (2009), and Moretti (2010) for surveys.

shock on rural locations.

The focus on rural locations is important, because agglomeration economies models imply that the impact of temporary shocks depends on the initial conditions. Specifically, once a core-periphery structure has emerged, changing the spatial configuration of economic activity requires a much larger shock than what is needed in an earlier stage of the process. My results suggest that while the major Finnish cities were well established by the mid-1940s, many rural locations were at the brink of becoming a local manufacturing center. The resettlement shock was sufficiently large to affect which of the potential equilibria materialized.

My findings also connect with the related literature on the interplay between growth, structural change and urbanization (e.g. Black and Henderson 1999, Caselli and Coleman 2001, Duranton 2007, Rossi-Hansberg and Wright 2007). In particular, I complement the recent work by Bleakley and Lin (2010) and Michaels, Rauch, and Redding (2010), who examine the evolution and determinants of population density in the United States. Unlike these studies, however, I study the long-term effects of a temporary shock.

Third, I add to the literature on the impact of large and sudden immigration flows (Card, 1990; Hunt, 1992; Carrington and de Lima, 1996; Friedberg, 2001). A potential problem of these studies is that immigration into one location may affect other locations through changes in production structure and native migration patterns (Borjas, Freeman, and Katz, 1997; Borjas, 2003). My results support the importance of such general equilibrium effects. However, in contrast to the previous empirical evidence, I find a *positive* effect on later in-migration and a strong impact on production structure.[2]

A limitation of this study is that the research design does not allow for distinguishing between the alternative microfoundations behind agglomer-

[2]Studies on the impact of immigration on native migration patterns include Frey (1995), Wright, Ellis, and Reibel (1997), Card and DiNardo (2000) and Card (2001). Hanson and Slaughter (2002), Lewis (2003) and Gandal, Hanson, and Slaughter (2004) examine the impact of immigration on production structure.

ation economies. During the past two decades, an active line of research has formalized Marshall's insight that proximity facilitates the flow of goods, people and ideas.[3] My results are consistent with each of these mechanisms, but do not measure their relative importance.

The rest of this paper is organized as follows. The next section sets up a model and discusses a hypothetical experiment that guides the empirical work. Section 3 discusses the resettlement in detail and Section 4 presents the data. Section 5 introduces the empirical strategy. Section 6 reports the results. Section 7 concludes.

2 A Test for Agglomeration Economies

I start by drafting a model that serves two purposes. First, it provides a simple framework that allows me to discuss some of the main insights of the theoretical literature on agglomeration. Second, it illustrates how exogenous migration flows can be used as a test for the presence of agglomeration economies.

2.1 Assumptions

Consider an economy that consists of J locations of equal area, each endowed with F_j units of an immobile factor and hosting N_j units of a mobile factor. Each location can produce two final goods, A and M. The mobile factor can be used for production in both sectors while the immobile factor is used only in sector A. For concreteness, I will call N workers, F the quality of land, A agriculture and M manufacturing.

Workers are homogeneous, are paid their marginal product, can move freely between sectors and locations, but cannot make collective migration

[3] Examples include, but are not limited to, Helsley and Strange (1990), Krugman (1991a,b), Acemoglu (1997), Helpman (1998), Fujita, Krugman, and Venables (1999), Glaeser (1999), Rotemberg and Saloner (2000) and Baldwin et al. (2003).

decisions. I abstract from labor supply decisions and assume that each worker provides one unit of labor and that utility equals wages. Final goods are traded in the international market, trade costs are zero and each location is too small to affect prices.

The location-level production function in agriculture is

$$Y_A = f\left(L_A, F\right)$$

where L_A is the amount of labor working in agriculture. Production in agriculture exhibits decreasing returns to scale: $f_L > 0$, $f_F > 0$, $f_{LL} < 0$ and $f_{FF} < 0$, where the subscripts refer to first and second derivates.

The location-level production function in manufacturing is

$$Y_M = \alpha\left(L_M\right) L_M$$

where L_M is the amount of labor working in manufacturing and $\alpha\left(L_M\right)$ is a productivity term.

I consider the implications of two alternative assumptions of the relationship between productivity and the size of the manufacturing sector. First, returns to scale may be constant, $\alpha'\left(L_M\right) = 0$, and thus productivity does not depend on L_M. Alternatively, productivity may increase with the size of the sector, $\alpha'\left(L_M\right) > 0$. For instance, a larger manufacturing sector could allow for more specialization in tasks or it might help workers to learn effectively from each other.[4]

[4] Adam Smith's discussion about the pin factory provides a classic example on the gains from task specialization. The learning hypothesis is typically attributed to Marshall (1890). See Duranton and Puga (2004) and Glaeser (2008) for modern expositions of the microfoundations of agglomeration economies.

2.2 The Regimes

Figure 1 illustrates the simple case where the economy consists of two locations that differ from each other only in that location b is endowed with a better quality of land than location a. In equilibrium, the supply and demand for labor must yield identical wages in both locations. Since labor supply is assumed to be perfectly inelastic, it corresponds to the size of the population. The labor demand curves—derived in the Web Appendix—are presented by the bold lines. They are downward sloping as long as the location fully specializes in the production of agricultural goods. Since location b has better land than location a, a given size of a labor force corresponds to higher wages in location b than a similar labor force in location a.

After the manufacturing sector has emerged, the shape of the labor demand curve depends on whether the manufacturing exhibits constant returns to scale (the left panel) or whether agglomeration economies are present (the right panel). In the former case, wages are fixed at $w_M = \alpha$. In the latter, there is a discontinuous jump at the point where the manufacturing sector emerges and an upward sloping labor demand curve thereafter.

The Natural Advantages Regime

I divide the parameter space of the model into three regimes that illustrate three broad explanations for the geography of economic activity. The 'natural advantages regime' corresponds to a small aggregate labor force. For example, if the aggregate population is $N_{0a} + N_{0b}$, the only allocation equalizing wages is such that N_{0a} workers live in location a and N_{0b} live in location b. To see this, note that if Δ workers move from a to b, wages at a increase to $w_{0a}^{'}$ and wages at b decrease to $w_{0b}^{'}$. As a result, workers migrate from b to a until the regional structure returns to its initial configuration. This example illustrates a general result: the natural advantages regime has a unique equilibrium, which is entirely determined by the endowments of the immobile factor.

The Random Growth Regime

The 'random growth regime' takes place when the aggregate population is large and returns to scale in manufacturing are constant. These parameter values lead to a large number of equilibria and the only role of the immobile factor is to determine the minimum population of each location. For example, if the aggregate population is larger than $N_{1a}+N_{1b}$ in the left panel of Figure 1, all configurations with $N_a > N_{1a}$ and $N_b > N_{1b}$ equalize wages and are thus an equilibrium. Note that they are not Pareto efficient. That is, if a sufficient number of workers move from a to b so that a fully specializes in agriculture, the resulting increase in wages would benefit workers at a, while workers at b would receive the same wage as before. However, higher wages at a would lead workers to migrate until wages are again equalized at w_M.

In this regime, changes in the labor supply are absorbed entirely through changes in the production structure. Thus it captures the intuition of the familiar Rybczynski (1955) theorem.[5] Furthermore, as temporary population shocks do not affect later population growth, this regime corresponds to the 'random growth' models following Simon (1955).[6]

The Agglomeration Economies Regime

In the third part of the parameter space, the aggregate population is large and productivity in manufacturing increases with the size of the sector. This 'agglomeration economies regime' has a multiple, but a smaller number of equilibria than the random growth regime. Within each equilibrium, natural advantages determine the regional structure. However, natural advantages

[5] The implications are not identical, however. In the Hecksher-Ohlin model underlying the Rybczynski theorem, growth in one factor leads to an absolute expansion in the product that uses that factor intensively and to an absolute contraction in the output of the product that uses the other factor intensively (as long as the location is not fully specialized). In the present model, land is used only in agriculture and thus the absolute size of agriculture is not affected by the size of the labor force.

[6] See Gabaix (1999) and Eeckhout (2004) for detailed discussion.

do not necessarily determine which equilibrium takes place.

Consider first the initial configuration $[N_{2a}, N_{2b}]$ in the right panel of Figure 1. In this case, wages are equalized and both locations fully specialize in agriculture. The equilibrium is stable to marginal shocks, but not to large shocks. That is, if one worker moves from a to b, wages at b will decrease, wages at a will increase and someone will move from b to a. However, if Δ workers move from a to b, a manufacturing sector emerges at b and wages jump to $w_{2b}^{'}$. Wages at a will also rise as there is now more land per worker. However, given the parameterization of this example, $w_{2b}^{'} > w_{2a}^{'}$. Thus workers keep on migrating in the same direction even after the initial migration flow. As a consequence, wages further increase both at a (because of decreasing returns) and at b (because of increasing returns). At the new equilibrium, only $N_{2a}^{''}$ workers stay at a while the population of b is $N_{2b}^{''}$.

The agglomeration economies regime illustrates some of the key insights of the 'new economic geography' literature following Krugman (1991b). Particularly, it shows how increasing returns to scale in manufacturing give rise to a regional structure consisting of a manufacturing core and an agricultural periphery. It also includes the 'history matters property'. That is, if Δ workers would have moved from b to a (instead of the other way around) at the initial configuration $[N_{2a}, N_{2b}]$, manufacturing sector would have emerged at a. However, if the starting point is $\left[N_{2a}^{''}, N_{2b}^{''}\right]$, the equilibrium will not change when Δ workers move from b to a.

A further interesting feature concerns the interaction between natural advantages and agglomeration economies. Note that *within* each equilibrium natural advantages determine the distribution of population. Furthermore, natural advantages can determine which core-periphery equilibrium takes place. To see this, note that if aggregate population grows steadily, the manufacturing sector first emerges at b. However, this is not an efficient outcome. Since location b is endowed with better land than location a, it has a comparative advantage in specializing in agriculture. Yet, the same comparative advantage led it to be the first to cross the threshold for setting up a manufacturing sector.

2.3 A Hypothetical Experiment

The simple model discussed above can incorporate all three leading explanations for the geographic concentration of economic activity. I will next ask whether a hypothetical experiment could allow researchers to empirically distinguish between these regimes. Particularly, I will think of an experiment that randomly allocates locations a and b into a treatment and control group and then moves Δ workers from some outside source into the treatment location.

Suppose that after the experiment, workers would start moving from the control to the treatment area. This finding would be consistent with the agglomeration economies regime and inconsistent with the natural advantages and random growth regimes. More precisely, the result would be consistent with two situations. First, the initial distribution of workers could be close to $[N_{2a}, N_{2b}]$ in Figure 1, in which case the experiment would push the treatment location over the threshold for the manufacturing sector to emerge. Alternatively, the initial configuration could be something like $\left[N''_{2a}, N''_{2b}\right]$ and, by chance, location b would be chosen as the treatment area.

Other findings would be more difficult to interpret. Suppose that after the experiment workers would start moving from the treatment to the control area. This result would be consistent with the natural advantages *and* the agglomeration economies regime. In the case of natural advantages, the experiment would decrease wages in the treatment area and thus workers would migrate to the control area until wages were again equalized.[7] In the agglom-

[7]This example illustrates the Borjas, Freeman, and Katz (1997) critique on the 'spatial correlations' approach used in much of the literature on the impact of immigration on native wages. Note that the experiment would decrease wages in both the treatment and control area. However, since wages would be equal in both area before and after the experiment, the spatial correlations approach would lead one to conclude that wages were

eration economies regime this result would occur if the treatment location was specialized in agriculture (e.g. location a when the initial conditions are $\left[N''_{2a}, N''_{2b}\right]$).

Similarly, a finding that the treatment location grew at the same rate as the control location would be hard to interpret. As long as migration costs are negligible, this finding would be consistent only with the random growth regime. However, as I discuss in detail in Section 6.4, the finding would be consistent with all regimes if migration costs were large. In this case, the impact of the hypothetical experiment on equilibrium wages would provide a test for each regime: a negative wage effect would be consistent only with the natural advantages regime, a zero effect with the random growth regime and a positive effect with the agglomeration economies regime.

While understanding the implications of this hypothetical experiment is useful for organizing thoughts, it is extremely unlikely that the experiment would be carried out in practice. However, sometimes historical episodes resemble the hypothetical experiment. I will next discuss whether the post-WWII population displacement in Finland qualifies as such a natural experiment.

3 The Finnish Resettlement Policy

At the beginning of World War II, Finland was a developing country, where half of the population worked in agriculture.[8] The war led Finland to cede over a 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

not affected at all.

[8]According to Maddison (2010), the Finnish GDP per capita was 3,589 International Geary-Khamis 1990 dollars in 1938. In comparison, Morocco, Algeria, Moldova, Jamaica, Egypt and Cuba had similar GDP per capita in 2008.

located in southeastern Finland, while two other ceded areas were located in the extremely sparsely populated northern parts of the country (see Figure 2).

The plan for resettling the evacuated population was designed in three pieces of legislation: the Rapid Resettlement Act, the Land Acquisition Act and the Settlement Plan. Those who had derived their principal income from agriculture in the ceded areas were entitled to receive cultivable land in the remaining parts of country. The displaced farmers were not able to choose their destination. Non-agrarian displaced persons received compensation for their lost property in the form of government bonds and were free to choose their destination areas.[9]

In total, roughly 250,000 hectares of existing cultivated land was used for resettlement and 150,000 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 on their farms—ended up providing roughly half of the cultivated fields. The land was purchased either on a voluntary basis or through expropriation using a progressive scale presented in Figure 3. 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 compensation in practice. That is, the expropriation did not inject cash into the affected municipalities. Furthermore, the affected areas were not targeted by any special regional policies. The allocation of land to displaced farmers was completed by the end of 1948.

The amount of land available for displaced farmers within the borders of a given municipality—and hence the number of displaced farmers allocated

[9]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 specific permission from the local housing board.

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 in 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. Given that 99 percent of the displaced farmers spoke Finnish as their mother tongue, very few of the displaced farmers received land from the Swedish-speaking parts of the country.

As I discuss in detail below, I will use these features of the resettlement policy to approximate the hypothetical experiment discussed in Section 2.3. The plausibility of this approach depends on the reasons why some locations were endowed with larger farms, more government-owned land or a larger Swedish-speaking population.

The origins of the identifying variation go back to the time when Finland formed the eastern part of Sweden.[10] 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, population expanded towards the east and north. A considerable number of migrants from Sweden also settled along the western and southern coasts. However, the vast area farther east and north remained a 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.

[10] Swedish rule started at around mid-12th century and ended in 1809 when Finland became part of Russia as an autonomous Grand Duchy. Since 1917, Finland has been an independent country. Throughout, 'Finland' refers to the area falling within the 1939 borders. For a more detailed description (in English) of the Finnish history, see e.g. Kirby (2006)

In short, the pattern of large farms in the southwest, government-owned land in the north and east, and Swedish-speaking settlements on the coasts was already present in the Middle Ages. This division faded over time, but there were still clear differences in the 1940s. Figure 2 illustrates these patterns. The bottom-right panel also presents the share of the displaced population 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 one-third increase in their populations. However, there was also large variation in the share of the displaced population between neighboring municipalities in the Finnish-speaking area.

Importantly, the historical economic advantage of the southwest virtually disappeared over time. One of the reasons was the rapid population growth and the end of the Little Ice Age, which pushed permanent settlement towards the east and north.[11] Another 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 city was moved eastwards from Turku to Helsinki. Furthermore, transportation technology improved substantially, particularly after the construction of an ambitious railroad network started in 1862. While the market area of St Petersburg disappeared with the Russian revolution and the consequent Finnish independence in 1917, Helsinki remained the capital city and the most important economic area of Finland.

[11] Between the mid-18th and mid-19th century, Finland experienced roughly 1.5 percent annual population growth. The Little Ice Age refers to the period of global cooling between the 16th and mid-19th century. While researchers do not agree on the exact timing of this period, there is a wide consensus that conditions for agriculture in northern and eastern parts of Finland improved substantially from the mid-18th century onwards.

4 Data

Most of my empirical analysis is based on a dataset that I have constructed using 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).[12] The data are further augmented with detailed local price indexes for 1980, an indicator variable for a municipality being connected to the railroad network in 1939 (as documented by historical engine driver timetables, see Kotavaara, Antikainen, and Rusanen forthcoming), and the number of displaced persons living in a municipality in 1948 (from an administrative report held at the National Archieves of Finland). In order to ensure that the spatial units remain stable over time, I 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 Web Appendix.

My second dataset was created by Statistics Finland and contains individual-level longitudinal information. The starting point is a sample of the original 1950 census forms, which were manually inserted into a database. These data were linked to the 1970 census data and 1971 tax records. Importantly, the 1950 census forms contained retrospective questions about the municipality of residence, socio-economic status and industry in 1939. Thus the data allows for distinguishing between the displaced persons and the local residents, who lived in the resettlement areas already in 1939. The original sample contains information on 411,629 people. I have access to a random sample of roughly a quarter of these data.

In the baseline analysis, I focus on those 349 rural municipalities that did not cede territory to the Soviet Union. Partly ceded municipalities are

[12]Municipalities in the baseline sample had a median land area of 417 square kilometers and a median population of 4,273 in the 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 the same year.

excluded, since consistent time series cannot be constructed for them. The main motivation for excluding cities is that the identification strategy relies on instruments that are relevant only for rural areas. In the Web Appendix, I show that the results are not sensitive to alternative sample selection rules.

Figure 4 plots the population growth rates between 1949 and 2000 on the growth rates between 1939 and 1949. It 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 war and its immediate aftermath, three quarters lost population during the next five decades. This decline was driven by emigration and, more importantly, migration into the larger metropolitan areas.

Migration from the countryside to larger cities had already begun before the war and it was particularly intensive in 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 a half in 2000.[13] Despite the relative decline, however, the number of people residing in the baseline sample area increased as the aggregate population grew. The study area had a population density of 7.7 persons per square kilometer in 1939 and 9.1 persons per square kilometer in 2000.

5 Empirical Strategy

In Section 2, I argued that self-reinforcing population growth would provide compelling evidence of the presence of agglomeration economies. However, establishing that population growth during one period *causes* population growth in a later period is challenging, because confounding factors could drive the population growth in both periods. I next discuss an empirical strategy that exploits the features of the Finnish resettlement policy to over-

[13] The calculation for 1930 excludes areas that were later ceded to the Soviet Union.

come this identification problem.

I will take a simple approach and estimate variants of

$$y_{jt} = \beta g_{jw} + \mathbf{X_j}\gamma + \epsilon_j \tag{1}$$

where y_{jt} is the outcome of interest in location j at period t, g_{jw} is the population growth rate between 1939 and 1949, $\mathbf{X_j}$ is a vector of observable characteristics measured before the war, and ϵ_j summarizes unobserved factors affecting the outcome. The parameter of interest is β.

The challenge in consistently estimating β is that ϵ_j could also affect wartime population growth. For example, if economically more viable locations received more migrants than locations with less potential for growth, the OLS estimates of β would be biased upwards. Alternatively, if the resettlement policy pushed people to areas that had low growth potential, the OLS estimates would be biased downwards.[14]

I address the issue in two ways. First, I control for pre-war observable characteristics and constant geographical characteristics. While these variables may not capture all factors affecting wartime population growth, conditioning on them should reduce the potential bias.

The main identification strategy, however, is to use an instrumental variables approach exploiting the three elements of the allocation policy discussed in Section 3. The instruments are the proportion of a 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 as predicted by the 1930 size distribution of privately owned land.[15]

[14] 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). For instance, the displaced persons were offered the option of receiving a "cold farm". These farms were located in eastern and northern Finland and had no cultivated land or buildings (Mead, 1951).

[15] I 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

5.1 First-Stage Estimates

The first column of Table 1 report the results of regressing municipality-level population growth between 1939 and 1949 on the instruments (i.e. the first-stage for the 2SLS estimates reported in the next section).[16] The results are in line with the resettlement policy. A larger stock of available agricultural land is positively associated with population growth rate during the resettlement period. Similarly, municipalities with a large proportion of Swedish-speaking people received fewer displaced persons and thus grew less. Together, the instruments explain roughly a sixth of the variance in the wartime population growth rate. The estimates are similar also in a specification controlling for pre-war municipality characteristics (second column). The F-statistics imply that the research design does not suffer from problems related to weak instruments.

The instrumental variables strategy hinges on the identifying assumptions that the instruments had no direct effect on post-war outcomes. As discussed in Section 3, this assumption seems plausible given that the identifying variation reflects economic conditions in the Middle Ages and that these factors had lost their relevance by mid-20th century. A data-driven way to assess the plausibility of this argument is to ask whether the instruments explain pre-war population growth. Results reported in Columns 3 and 4 of Table 1 suggest that they do not. The only statistically significant association is between the availability of privately owned land and pre-war population in a specification where I do not control for 1930 characteristics. I return to the potential implications of this association in Section 6.2.

---

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 the municipality's 1939 population.

[16]The first-stage estimates using each instrument individually are similar to the estimates presented in Table 1 (see Table A1 in the Web Appendix).

6 Results

6.1 Population Growth

The impact of a population shock on later population growth provides arguably the most powerful test for the presence of agglomeration economies. As discussed in detail in Section 2.3, if immobile factors determine the spatial distribution of economic activity, a positive spurious shock in one period will have a negative impact on the population growth of the next period. In the random growth models, one-off population shocks will not affect later population growth at all. In a model with agglomeration economies, population growth during one period may have a negative or a positive effect on later population growth depending on the initial conditions.

Table 2 reports estimates for the impact of the resettlement shock on the population growth rate in the post-war period. Each estimate stems from separate regression, which differ in the length of the post-war period studied, the estimation method used and the inclusion of control variables. For example, the first column of panel A reports OLS (first row) and 2SLS (second row) estimates from regressing population growth rate between 1949 and 1950 on the population growth rate between 1939 and 1949 and a constant. Similarly, the sixth column presents the estimates from regressing population growth rate between 1949 and 2000 on population growth rate between 1939 and 1949 and a constant. Panel B reports corresponding estimates after controlling for pre-war municipality characteristics and geographical indicators.[17]

The results suggest that the resettlement shocks increased later popula-

[17] The control variables are the population growth rate between 1930 and 1939, the share of the labor force working in the primary sector in 1930, natural logarithm of the mean taxable income per capita in 1939, natural logarithm of the population density in 1939, natural logarithm of the distance weighted sum of the population of all municipalities in 1939 (using the inverse of Euclidean kilometer distance as weights), an indicator for sharing a border with a city, an indicator for being connected to the railroad network in 1939, longitude and latitude.

tion growth. According to the 2SLS point estimates reported in the sixth column of panel B, an exogenous migration flow increasing a municipality's population by 10 percent during the war caused an additional 17 percent population growth during the next five decades. All estimates are positive and statistically highly significant.

6.2 Robustness Checks

The finding that the resettlement shock increased later population growth is consistent with the presence of agglomeration economies and inconsistent with the natural advantages and random growth models. However, concluding that the results follow from agglomeration economies could be mistaken for two reasons. First, the estimates might be biased upwards. Second, a causal relationship between the wartime and post-war population growth rates could reflect some other mechanism than agglomeration economies. This subsection examines these possibilities in detail and argues that agglomeration economies remain the most likely explanation for the results discussed above.

Causality

A concern that the OLS estimates may be biased upwards is certainly reasonable. Any unobserved factor that affects population growth in the same direction during the resettlement period and the post-war period would lead to such a bias. While controlling for pre-war characteristics should help, controlling for all relevant factors may be beyond the scope of the data. Thus the OLS estimates alone would not provide compelling evidence on causality.

The instrumental variables estimates would be biased upwards if the land instruments had a positive direct effect or if the share of the Swedish-speaking population had a negative direct effect on the post-war population growth (see the Web Appendix for detailed discussion). Note that the data do not provide support for the availability of land having a positive impact on population growth. If anything, the estimates for pre-war population growth presented in columns 3 and 4 of Table 1 suggest that the land instruments were negatively associated with population growth and would thus bias the second-stage estimates downwards. However, the point estimates for the share of Swedish-speaking population are negative although statistically insignificant.

Table 3 reports the 2SLS estimates for the population growth rate between 1949 and 2000 using the instruments separately. Comparison of these estimates is informative as each instrument affects a very different area. Particularly, large privately owned farms were mostly located in the prosperous southwest, while government owned land was concentrated in the east and the north (see Figure 2).[18] It seems unlikely that *both* instruments would have a direct positive effect on post-war population growth. Yet, as shown in columns (2) and (3) of Table 3, they yield almost identical point estimates when used individually. On the other hand, the point estimates using only the share of Swedish speaking population as an instrument are positive, but smaller in magnitude and statistically insignificant. However, as shown in column (6), the null that all estimates are the same cannot be rejected. These findings—together with a range of further robustness checks and falsification exercises reported in the Web Appendix—strongly support the causal interpretation of the instrumental variables results.[19]

[18] 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 proportion of Swedish-speaking people and privately owned land (publicly owned land) is 0.01 (−0.06).

[19] In the Web Appendix, I show that the exclusion restriction would have to be violated by large magnitude in order to change the results qualitatively, I present results using alternative sample areas and subsamples where I gradually exclude the most influential observations (outliers), and I report the results using various alternative modes of inference including Conley's (1999) spatial GMM estimates and standard errors. All approaches yield qualitatively similar results.

Alternative Mechanisms

Consider next whether a causal impact of the resettlement shock could have be driven by other mechanisms than agglomeration economies. The simplest possibility is that the displaced population might have had higher fertility or lower mortality rates. In order to assess this channel, Table 4 examines the differences between the displaced and non-displaced populations using the individual-level Census data for persons who remained of working age between 1939 and 1950. Columns 1 and 2 report means for the entire samples of the displaced and non-displaced persons. Columns 3 and 4 report similar means for a sample consisting of persons living in the resettlement area before the war and of the displaced persons who lived in rural areas before the war. The comparison suggests that the fertility rates among the displaced population were slightly lower and mortality rates slightly higher than among the non-displaced population. Thus the post-war population growth appears to be caused by migration.[20]

The results would also be consistent with the resettlement increasing the stock of capital in the affected locations. However, as discussed in Section 3, the resettlement policy did not inject cash to the affected locations. Furthermore, since most of the capital owned by the displaced persons was left in the ceded area, the resettlement decreased capital-labor ratios. Thus improvements in the relative magnitude of physical capital do not explain the results. Furthermore, Table 4 does not provide any indication for the displaced population having brought scarce human capital to the resettlement areas. Particularly, there were no statistically significant differences in formal education, in the propensity to be an entrepreneur or in being a white-collar worker between the displaced persons and the locals. In fact, the displaced persons were slightly less likely to participate in the formal labor market in 1939 than the non-displaced persons.

[20]Further evidence using post-war migration flow data support this conclusion (Section B.6 of the Web Appendix).

6.3 Production Structure

The impact of the resettlement shock on post-war population growth rate suggests that agglomeration economies were important in the resettlement area. I will next ask whether also the impacts on the production structure are in line with the presence of agglomeration economies.

Recall that in the simple model of Section 2, a population shock would give rise to self-reinforcing population growth only if the affected location was close or above the threshold where a manufacturing sector emerges. Unfortunately, I cannot determine which municipalities were just below this threshold before experiencing the resettlement shock as the threshold is a function of location's endowments of the immobile factors—most of which are not observed in the data. Thus, I will take an approximate approach and regress the post-war population growth on wartime population growth, the share of the labor force working in the non-primary sector before the war and their interaction.

Table 5 reports the results. The estimates suggest that the impact of the resettlement shock was larger among municipalities that already had some non-primary production before the war. For example, a median wartime population growth rate (19 percent) would increase the post-war population growth rate by 15 percent at the first quartile (7 percent of the labor force working in the non-primary sector in 1939) and by 25 percent at the third quartile (22 percent at the non-primary sector). I interpret these results to be broadly in line with the model's predictions.

The model also predicts that the manufacturing sector should drive the growth of the labor force. Table 6 examines this hypothesis by regressing the post-war growth rate of the labor force working in the primary (first row) and the non-primary sector (second row) on the wartime population growth rate. As expected, the growth of the labor force occurs entirely at the non-primary sector.

6.4 Extensions

Thus far the analysis has been motivated by a highly stylized model building on the assumption of wage equalization across locations. In a richer model, differences in housing prices or local amenities could allow equilibrium wages to differ across regions (Rosen 1979, Roback 1982). I will next discuss the extent to which I can assess the importance of these channels in the case of the post-WWII Finnish population resettlement.

As in Moretti (2010), suppose that worker i in location j at time t has an indirect utility function

$$U_{ijt} = w_{jt} - r_{jt} + A_{jt} + e_{ijt} \tag{2}$$

where w_{jt} is the nominal wage rate, r_{jt} is the housing cost, A_{jt} is a measure of local amenities and e_{ijt} represents individual-level idiosyncratic preferences for location j. A full model would also include the dynamics of the housing market and the local amenities and define the distribution of the location-specific idiosyncratic preferences. In the interest of keeping the paper short, however, I discuss the implications informally.[21]

Wages

I examine the impact of the resettlement shock on long-term nominal wages using the linked census and tax register data. I restrict the analysis to individuals born between 1905 and 1939, who resided in the settlement area in 1939. I use log annual taxable income in 1971—the first period for which income data is available—as the dependent variable and the population growth rate between 1939 and 1949 in the municipality where the person lived in 1939 as the treatment variable. In addition, some specifications control for pre-war characteristics of the individual and the characteristics of his pre-war

[21] See Moretti (2010) for formal discussion.

municipality of residence.[22]

Panel A of Table 7 reports the results. The OLS estimates show a positive association between the wartime population growth rate and the income of local residents a quarter of a century later. The 2SLS estimates suggest that the association is causal. According to the 2SLS point estimates, ten percentage points increase in the wartime population growth rate increased the long-term income of the locals by roughly nine percent. Given that the average wartime growth rate in the baseline sample was 22 percent, these estimates imply that the resettlement had a substantial effect on local wages.

Interestingly, the 2SLS estimates are larger than the OLS estimates. There are at least two potential explanations for this finding. First, it may reflect the fact that the 2SLS estimator identifies a weighted average of local average treatment effects (Angrist and Imbens, 1995). That is, these estimates primarily capture the impacts on people living in municipalities whose growth rates were most affected by the resettlement policy. It is possible that wages in these locations were particularly responsive to the size of the labor force. Second, the OLS estimates may be biased downwards. This would occur, for example, if the wartime population growth was larger in locations where low housing prices or high local amenities compensated for (permanently) low nominal wages. Unfortunately, available data do not allow me to investigate these possibilities in detail.

Prices

The Rosen-Roback type of spatial equilibrium models suggest that the positive impact of the resettlement shock on nominal wages should be offset by an increase in local prices or a decrease in the value of local amenities. I examine

[22]The individual-level control variables are: age, age squared, gender, an indicator for speaking Swedish as one's mother tongue, six categories for socioeconomic status (entrepreneur, white-collar worker, blue-collar worker, assisting family member, out of labor force) and four categories for the sector of employment (agriculture, manufacturing, construction and services). The municipality level control variables are listed in footnote 17.

these channels using local price index data collected in 1980.[23] These data contain information on quality-adjusted housing costs, commodity prices and the travel cost associated with purchasing the commodities.

Panel B of Table 7 reports the results for housing prices. The OLS estimates suggest a positive association between the population growth rate in 1939–1949 and housing prices in 1980. However, the magnitude of this correlation is small and the 2SLS estimates are not statistically significant. Hence, the estimates imply that the long-term supply of housing was sufficiently elastic to accommodate the growing population. Of course, the finding is not surprising given the abundance of land in the affected locations and the fact that housing prices are measured 35 years after the war ended. Particularly, the results do not rule out the possibility that the resettlement shock would have increased short-term housing prices. Nevertheless, the results suggest that the resettlement shock had a permanent positive effect on real wages.

Amenities and Migration Costs

The indirect utility function (2) provides two possible explanations for why the resettlement could have permanently increased real wages. First, it could have decreased the quality of local amenities. Second, location-specific idiosyncratic preferences could create migration costs, which would prevent wage equalizing migration flows from taking place.

Consider first the impact of the resettlement shock on local amenities. Clearly, the displaced persons did not alter factors such as the climate. Furthermore, it seems unlikely that congestion would be an important problem in the affected locations as I examine very sparsely populated areas. In fact, the resettlement shock could have *improved* local consumption amenities.

[23] These data were colleceted to determine cost-of-living adjustments in centralized wage negotiations. The study collected data for rents and characteristics of 325,013 housing units, management expenses of 7,052 owner-occupied housing units, prices for 9,933 apartments and 4,466 detached houses sold in 1980, and 235,155 commodity prices from 34,503 shops. The cost of collecting the data was considerable and no local price indexes have been constructed in Finland after 1980.

Particularly, in the 'new economic geography' models building on monopolistic competition and non-trivial transport costs, population growth would increase the variety of brands available in a location (Krugman, 1991b; Fujita, Krugman, and Venables, 1999; Baldwin et al., 2003).[24] Thus consumers would benefit from lower prices (due to increased competition) and from the availability of a larger variety of products.

Panel C of Table 7 examines these effects using the local price index data for 1980. In the first row, the dependent variable is the local commodity price index for 317 items. For each commodity, prices were collected for the brands that had the largest markets share in the national market. If a brand was not available in the municipality, the price at the nearest location where the commodity was available was used. The estimates suggest that this index was not affected by the wartime population growth.

In the next row, the dependent variable is the travel costs for purchasing the basket of goods used for constructing the commodity price index. The travel cost estimate was based on the distance to the nearest shop selling each good and the typical frequencies of purchases by a commodity. The estimates are negative and statistically significant. This implies that shops located in municipalities that grew fast during the war were offering a wider variety of brands in 1980.

These results suggest that while the resettlement shock created migration responses, the migration flows were not sufficient to equalize wages net of local prices and amenities. This finding implies that migration costs were non-trivial. In the context of indirect utility function (2), migration costs would correspond to the location-specific idiosyncratic preferences. Other potential sources include, but are not limited to, incomplete information about the labor market opportunities and the direct monetary cost of moving. However, detailed examination of these mechanisms is beyond the scope of this paper.

[24] See Helpman (1998) for a discussion of immobile goods in a similar framework.

7 Conclusions

I have examined the long-term impact of resettling more than a tenth of the Finnish population after World War II. This historical episode allows for constructing plausible instrumental variables that can be used to estimate the causal impact of labor supply shocks on later outcomes. In line with an agglomeration economies model—and in contrast to other popular models explaining the spatial distribution of economic activity—I find that the resettlement shock increased population growth, industrialization and wages.

Most of the previous empirical work on agglomeration has focused on cities in developed countries. While these cities are clearly important engines of growth and innovation, they host a relatively small share of world's population. Thus one might hypothesize that even if agglomeration economies are important in Silicon Valley or Manhattan, they might be irrelevant in the areas where most people reside.

The results reported in this paper suggest otherwise. I focus on an area where population density remains below ten inhabitants per square kilometer. At the time of the resettlement, 80 percent of the population in the baseline sample area was working in agriculture and the Finnish GDP per capita was comparable to today's middle-income developing countries. Yet, I find robust evidence on self-reinforcing growth. Thus I interpret these findings to support the general importance of agglomeration economies.

References

Acemoglu, D. 1997. “Training and Innovation in an Imperfect Labour Market.” *Review of Economic Studies* 64 (3):445–464.

Acemoglu, D., T.A. Hassan, and J.A. Robinson. forthcoming. “Social Structure and Development: A Legacy of the Holocaust in Russia.” *Quarterly Journal of Economics* .

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

Angrist, J.D. and G.W. 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.

Arzaghi, M. and J.V. Henderson. 2008. “Networking off Madison Avenue.” *Review of Economic Studies* 75 (4):1011–1038.

Baldwin, R.E., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud. 2003. *Economic Geography and Public Policy*. Princeton University Press.

Black, D. and V. Henderson. 1999. “A Theory of Urban Growth.” *Journal of Political Economy* 107 (2):252–284.

Bleakley, H. and J. Lin. 2010. “Portage: Path Dependence and Increasing Returns in US History.” NBER WP No. 16314.

Borjas, G. 2003. “The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market.” *Quarterly Journal of Economics* 118 (4):1335–1374.

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.

———. 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. 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.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 43 (2):245–257.

Carrington, William J. and Pedro 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.

Caselli, F. and W.J. Coleman. 2001. "The US Structural Transformation and Regional Convergence: A Reinterpretation." *Journal of Political Economy* 109 (3):584–616.

Ciccone, A. and R.E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86 (1):54–70.

Combes, P.P., G. Duranton, L. Gobillon, D. Puga, and S. Roux. 2009. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." CEPR WP 7191.

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

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

———. 2008. " Search for Multiple Equilibria in Urban Industrial Structure." *Journal of Regional Science* 48 (1):29–65.

Duranton, G. 2007. "Urban evolutions: The fast, the slow, and the still." *the American Economic Review* 97 (1):197–221.

Duranton, G. and H.G. Overman. 2005. "Testing for localization using micro-geographic data." *Review of economic studies* 72 (4):1077–1106.

Duranton, G. and D. Puga. 2004. "Micro-foundations of urban agglomeration economies." *Handbook of regional and urban economics* 4:2063–2117.

Eeckhout, J. 2004. "Gibrat's Law for (all) Cities." *American Economic Review* :1429–1451.

Ellison, G. and E.L. Glaeser. 1997. "Geographic Concentration in US Manufacturing Industries: a Dartboard Approach." *Journal of Political Economy* 105 (5):889–927.

———. 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review* 89 (2):311–316.

Ellison, G., E.L. Glaeser, and W.R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100 (3):1195–1213.

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

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

Fujita, M., P.R. Krugman, and A. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press.

Gabaix, X. 1999. "Zipf's Law For Cities: An Explanation." *Quarterly journal of Economics* 114 (3):739–767.

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

Glaeser, E.L. 1999. "Learning in Cities." *Journal of Urban Economics* 46 (2):254–277.

———. 2008. *Cities, Agglomeration, and Spatial Equilibrium*. Oxford University Press.

Glaeser, E.L. and J.D. Gottlieb. 2009. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature* 47 (4):983–1028.

Glaeser, E.L. and M.G. Resseger. 2010. "The Complementarity Between Cities and Skills." *Journal of Regional Science* 50 (1):221–244.

Greenstone, M., R. Hornbeck, and E. Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118 (3):536–598.

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

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

Hanson, G.H. and C. Xiang. 2004. "The Home-Market Effect and Bilateral Trade Patterns." *American Economic Review* :1108–1129.

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.

———. 2004b. "The Empirics of Agglomeration and Trade." In *Handbook of Regional and Urban Economics*, vol. 4, edited by J. V. Henderson and J. F. Thisse, chap. 59. Elsevier, 2609–2669.

Helpman, Elhanan. 1998. "The Size of Regions." In *Topics in Public Economics: Theoretical and Applied Analysis*, edited by David. Pines, Efraim. Sadka, and Itzhak. Zilcha. Cambridge: Cambridge University Press, 33–54.

Helsley, R.W. and W.C. Strange. 1990. "Matching and Agglomeration Economies in a System of Cities." *Regional Science and Urban Economics* 20 (2):189–212.

Hornbeck, R. 2009. "The Enduring Impact of the American Dust Bowl: Short and Long-run Adjustments to Environmental Catastrophe." NBER WP No. 15605.

Hunt, Jennifer. 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: Cambridge University Press.

Kotavaara, O., H. Antikainen, and J. Rusanen. forthcoming. "Accessibility concentrating population: a GIS analysis relating the history of Finnish urbanization to the accessibility by road and railway networks in 1880–1970." *Journal of Interdisciplinary History* .

Krugman, P. 1991a. *Geography and Trade*. Cambridge, MA: MIT Press.

———. 1991b. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99 (3):483.

Laitinen, E. 1995. "Vuoden 1945 maanhankintalain synty." In *Rintamalta raiviolle. Sodanjälkeinen asutustoiminta 50 vuotta*, edited by E. Laitinen. Jyväskylä: Atena Kustannus, 52–138.

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

Maddison, A. 2010. "Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD." Dataset available at http://www.ggdc.net/maddison/.

Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.

Mead, WR. 1951. "The Cold Farm in Finland: Resettlement of Finland's Displaced Farmers." *Geographical Review* :529–543.

Michaels, G., F. Rauch, and S.J. Redding. 2010. "Urbanization and Structural Transformation." Mimeo. Princeton and LSE.

Miguel, E. and G. Roland. forthcoming. "The Long Run Impact of Bombing Vietnam." *Journal of Development Economics* .

Moretti, E. 2010. "Local labor markets." NBER WP No. w15947.

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

Partridge, M.D., D.S. Rickman, K. Ali, and M.R. Olfert. 2009. "Agglomeration Spillovers and Wage and Housing Cost Gradients across the Urban Hierarchy." *Journal of International Economics* 78 (1):126–140.

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

Rappaport, J. and J.D. Sachs. 2003. "The United States as a Coastal Nation." *Journal of Economic Growth* 8 (1):5–46.

Redding, S.J. and D.M. 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. forthcoming. "History and Industry Location: Evidence from German Airports." *Review of Economics and Statistics* .

Roback, J. 1982. "Wages, rents, and the quality of life." *The Journal of Political Economy* 90 (6):1257–1278.

Rosen, S. 1979. "Wage-based indexes of urban quality of life." *Current issues in urban economics* 3.

Rosenthal, Stuart S. and William C. Strange. 2001. "The Determinants of Agglomeration." *Journal of Urban Economics* 50 (2):191 – 229.

———. 2003. "Geography, industrial organization, and agglomeration." *Review of Economics and Statistics* 85 (2):377–393.

———. 2004. "Evidence on the nature and sources of agglomeration economies." In *Handbook of Regional and Urban Economics*, vol. 4, edited by J. Vernon Henderson and Jacques-Franï¿œois Thisse. Elsevier, 2119 – 2171.

———. 2008. "The attenuation of human capital spillovers." *Journal of Urban Economics* 64 (2):373 – 389.

Rossi-Hansberg, E. and MLJ Wright. 2007. "Urban Structure and Economic Growth." *Review of Economic Studies* 74:597–624.

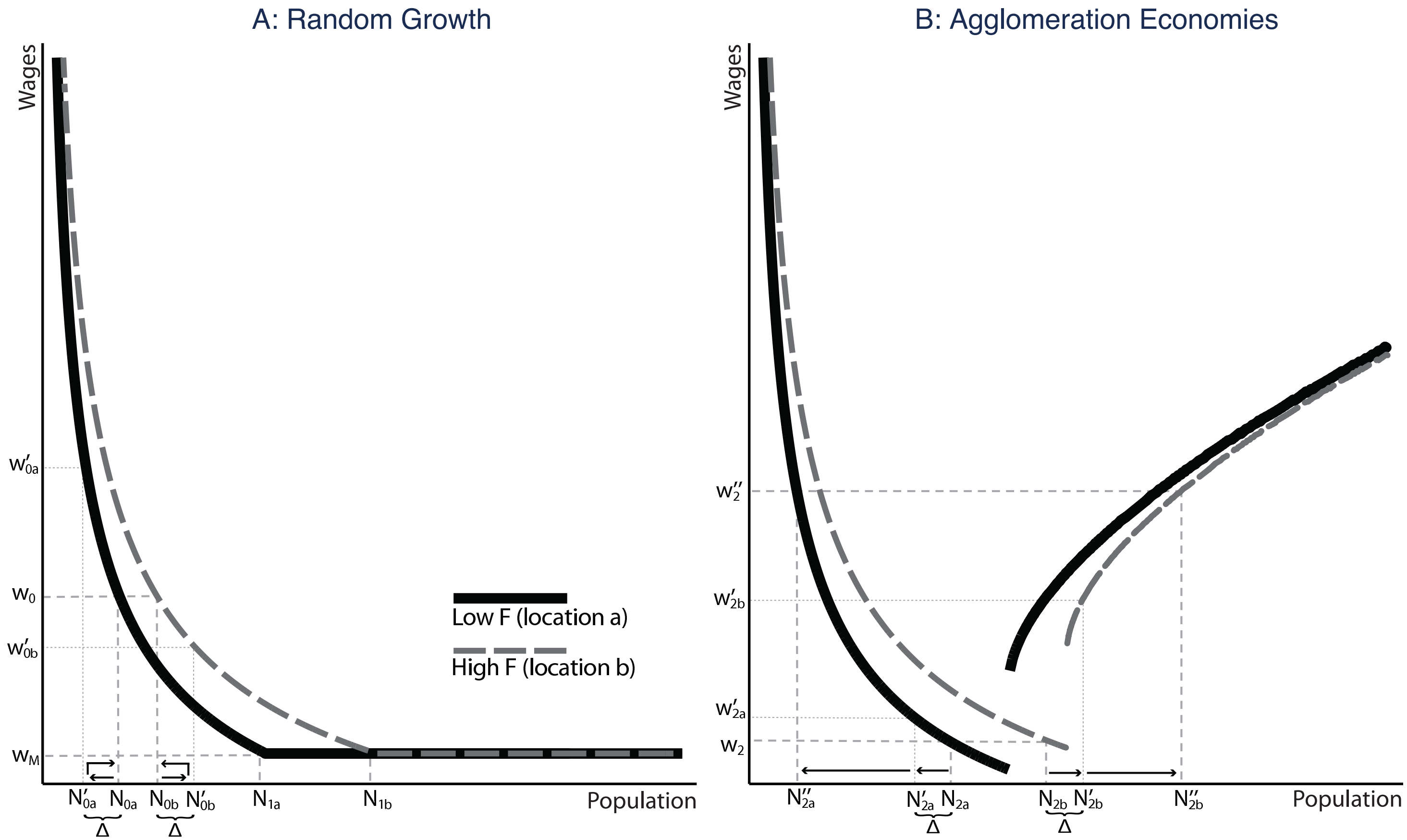
Rotemberg, J.J. and G. Saloner. 2000. "Competition and Human Capital Accumulation: a Theory of Interregional Specialization and Trade." *Regional Science and Urban Economics* 30 (4):373–404.

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

Simon, H.A. 1955. "On a Class of Skew Distribution Functions." *Biometrika* 42 (3-4):425–440.

Wright, R.A., 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: Equilibrium in a Two Location Example



Note: The bold lines plot the marginal product of labor within location (y-axis) as a function of its population (x-axis). Labor supply is assumed perfectly inelastic and thus corresponds to the size of the population. Location b is endowed with better quality of land than location a. The downward sloping part of the labor demand curves corresponds to the 'natural advantages' regime where the aggregate population is too small to support a manufacturing sector. The flat part of the labor demand curve in the left panel corresponds to the 'random growth' regime, where where the manufacturing sector is present and exhibits constant returns to scale. The upward sloping part of the labor demand curve in the right panel corresponds to the 'agglomeration economies' regime, where the manufacturing sector is present and exhibits external economies of scale.

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

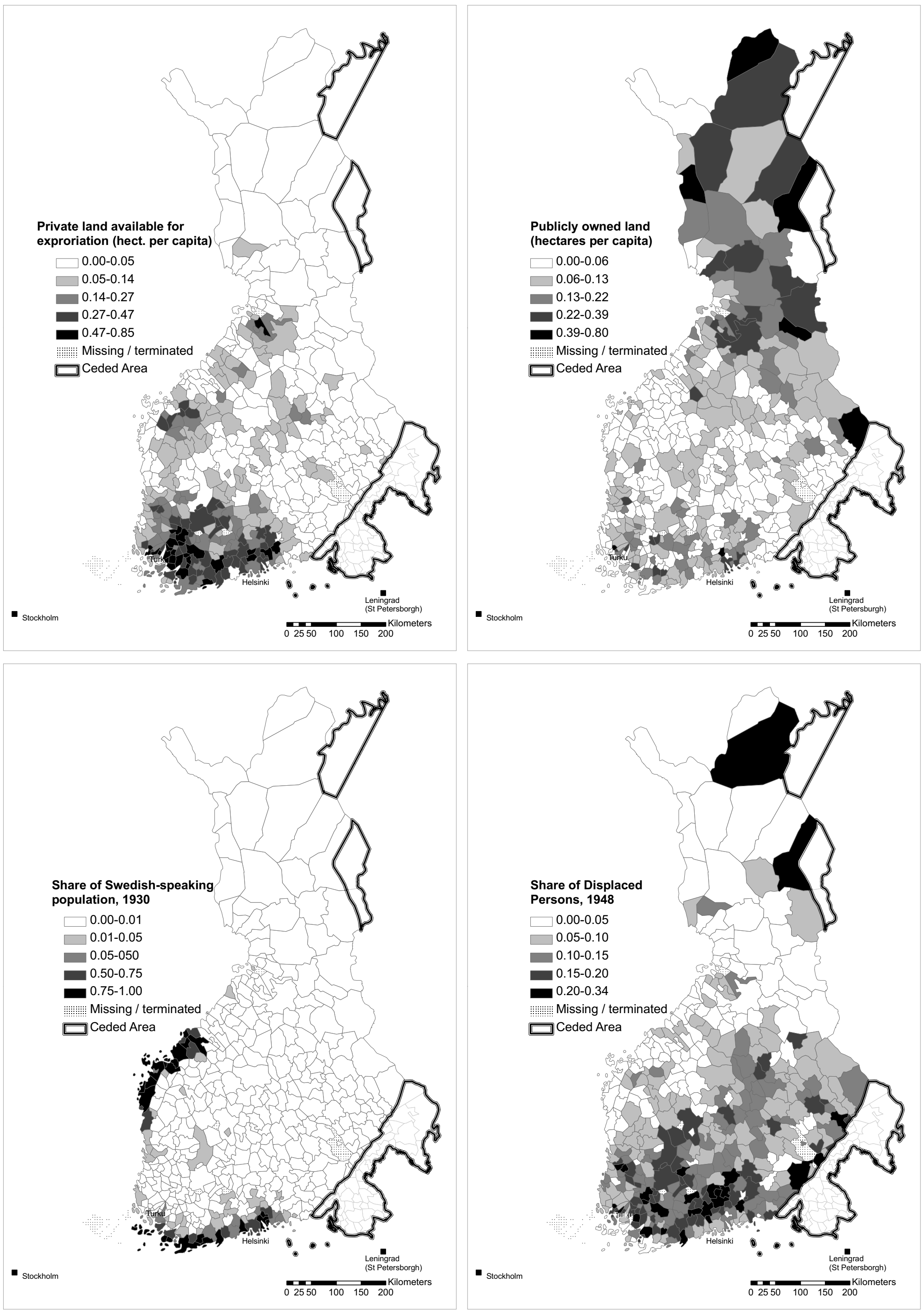
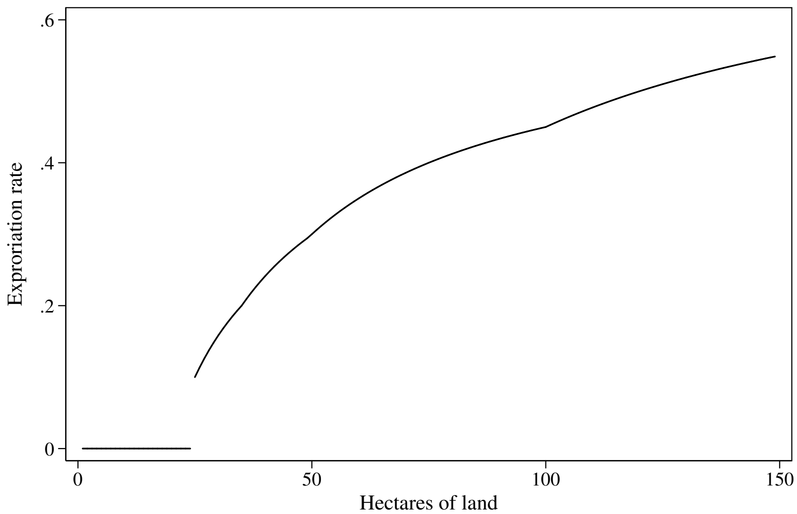
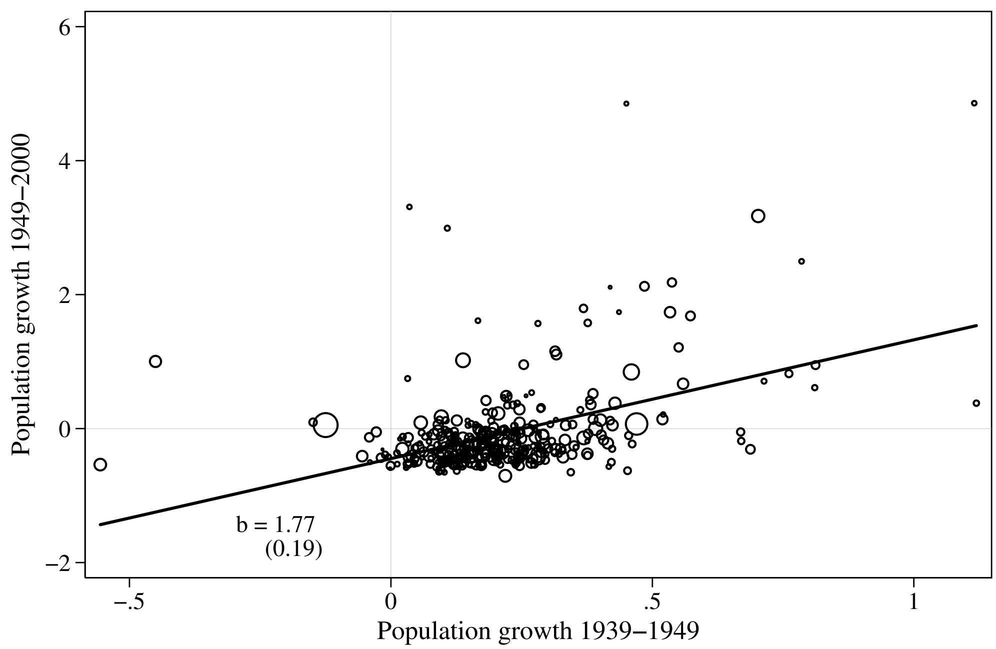


Figure 3: Expropriation Rate for Privately Owned Agricultural Land



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 4: Wartime and Post-War Population Growth



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

Table 1: The First-Stage and a Falsification Exercise

	First-Stage (population growth rate 1939–1949)		Falsification Exercise (population growth rate 1930–1939)	
	(1)	(2)	(3)	(4)
Hectares of expropriable land per capita (1930)	0.24	0.24	-0.18	-0.03
	(0.05)	(0.06)	(0.07)	(0.08)
Hectares of publicly owned land per capita (1940)	0.25	0.12	0.00	-0.18
	(0.10)	(0.11)	(0.14)	(0.16)
Share of Swedish-speaking population (1930)	-0.19	-0.16	-0.04	-0.05
	(0.04)	(0.04)	(0.05)	(0.05)
Control variables	no	yes	no	yes
F-statistic for the instruments	21.5	17.7	2.8	1.0
Partial R^2	0.16	0.14	0.02	0.01

Note: OLS estimates and standard errors (in parentheses). Sample: 349 rural municipalities. Control variables for column 2: population growth rate between 1930 and 1939, the share of the labor force working in the primary sector in 1930, natural logarithm of the mean taxable income per capita in 1939, natural logarithm of the population density in 1939, natural logarithm of the distance weighted sum of the population of all municipalities in 1939 (using the inverse of Euclidean kilometer distance as weights), an indicator for sharing a border with a city, an indicator for being connected to the railroad network in 1939, longitude and latitude. Control variables for column 4: the share of the labor force in the primary sector 1930, natural logarithm of population density in 1930, natural logarithm of distance weighted sum of the population of all municipalities in 1930, an indicator for sharing a border with a city, longitude, and latitude.

Table 2: The Impact on Post-War Population Growth Rate

	Dependent variable: Population Growth Rate 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	0.43	0.76	1.09	1.33	1.51
	(0.01)	(0.06)	(0.08)	(0.13)	(0.17)	(0.20)
2SLS	0.11	0.36	0.59	0.93	1.32	1.65
	(0.03)	(0.15)	(0.22)	(0.34)	(0.45)	(0.54)

Note: OLS and 2SLS estimates for the population growth between 1939 and 1949 and standard errors (in parentheses). Each estimate stems from separate regression. Sample, instruments and control variables: see Table 1.

Table 3: Main Estimates by Instrument

		2SLS				Sargan-
	OLS	Inst. 1	Inst. 2	Inst. 3	All	Hansen
	(1)	(2)	(3)	(4)	(5)	(6)
Without control variables	1.77	1.84	1.79	0.49	1.28	1.82
	(0.19)	(0.70)	(1.03)	(0.80)	(0.49)	[0.40]
With control variables	1.51	2.14	2.47	0.97	1.65	1.85
	(0.20)	(0.69)	(1.25)	(0.74)	(0.54)	[0.40]

Note: OLS and 2SLS estimates for the population growth between 1939 and 1949 and standard errors (in parentheses). Each estimate stems from separate regression. Dependent variable: Population growth rate in 1949–2000. Instrument 1: Hectares of expropriable land per capita (1930). Instrument 2: Hectares of publicly owned landper capita (1940). Instrument 3: Share of Swedish-speaking population (1930). Sample: 349 rural municipalities. Column (6) reports the Sargan-Hansen test-statistics and its p-value [in brackets]. Control variables: see Table 1.

Table 4: Comparisons of the Displaced and the Non-Displaced Populations

	Full Sample		Resettlement Sample	
	Non-Displaced (1)	Displaced (2)	Non-Displaced (3)	Displaced (4)
Female	0.55	0.55	0.53	0.55
#Children, 1950	1.39	1.37	1.54	1.44
#Children, 1970	0.91	0.79	1.04	0.85
Alive, 2002	0.32	0.30	0.32	0.30
Native tongue Swedish	0.10	0.01	0.06	0.00
Post-primary education, 1950	0.09	0.09	0.04	0.05
Socio-Economic Status, 1939				
Entrepeneur (non-agriculture)	0.04	0.03	0.03	0.03
Entrepeneur (farmer)	0.10	0.11	0.14	0.15
White Collar	0.10	0.11	0.06	0.07
Blue Collar	0.32	0.25	0.27	0.21
Assisting Family Member	0.13	0.16	0.19	0.21
Out of Labor Force	0.31	0.34	0.31	0.34
Sector of Employment, 1939				
Agriculture	0.31	0.30	0.44	0.41
Manufacturing	0.14	0.10	0.09	0.07
Construction	0.04	0.04	0.03	0.04
Services	0.17	0.19	0.11	0.13
Unknown	0.34	0.37	0.33	0.35
Observations	3,426	31,395	2,485	19,762

Note: Means for displaced and non-displaced persons. Sample: Cohorts born between 1885 and 1924. The sample for columns 3 and 4 includes only persons who lived in the baseline sample area in 1939 and displaced person who lived in rural areas before the war.

Table 5: Interactions with the Initial Production Structure

	OLS		2SLS	
	(1)	(2)	(3)	(4)
Population Growth Rate between 1939 and 1949	0.81	0.80	0.14	0.58
	(0.31)	(0.33)	(0.57)	(0.66)
Labor Force Share in the Non-Primary Sector, 1930	0.69	0.85	-0.78	-0.59
	(0.34)	(0.39)	(0.58)	(0.90)
Interaction	2.51	2.53	7.25	6.52
	(0.94)	(0.93)	(1.87)	(2.46)
Control variables	no	yes	no	yes

Note: OLS and 2SLS estimates and standard errors (in parentheses). Dependent variable: Population growth rate between 1949 and 2000. Instruments: see Table 2. Control variables: population growth between 1930 and 1939, log taxable income per capita in 1939, log 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 the railroad network in 1939.

Table 6: The Impact on the Production Structure, 1950–2000

	OLS		2SLS	
	(1)	(2)	(3)	(4)
Primary Sector	0.02	0.01	-0.01	0.00
	(0.01)	(0.01)	(0.02)	(0.02)
Non-Primary Sector	0.21	0.23	0.41	0.62
	(0.08)	(0.08)	(0.19)	(0.24)
Control variables	no	yes	no	yes

Note: OLS and 2SLS estimates for the population growth rate between 1939 and 1949 and standard errors (in parentheses). Dependent variables: percentage change in the number of individuals working in primary (first row) and non-primary (second row) sector between 1950 and 2000. Instruments: see Table 2. Control variables: population growth between 1930 and 1939, taxable income per capita in 1939, the share of the labor force in the primary sector in 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 the railroad network in 1939.

Table 7: The Impact on Wages, Prices and Local Amenities

	OLS		2SLS	
	(1)	(2)	(3)	(4)
A: Nominal wages (1971)				
log Annual Taxable Income	0.19	0.14	0.67	0.89
	(0.06)	(0.06)	(0.18)	(0.23)
B: Housing Market (1980)				
Housing price index	0.07	0.06	-0.06	-0.06
	(0.02)	(0.02)	(0.05)	(0.05)
C: Consumption amenities (1980)				
Commodity price index	0.00	0.00	-0.03	-0.02
	(0.01)	(0.01)	(0.01)	(0.02)
Travel cost associated with aquiring the CPI basket	-0.13	-0.06	-0.27	-0.16
	(0.04)	(0.03)	(0.09)	(0.08)
Control variables	no	yes	no	yes

Note: OLS and 2SLS estimates for the population growth rate between 1939 and 1949. Sample (panel A): 17,060 persons born between 1905 and 1939, who lived in the future resettlement area in 1939. Sample (panels B and C): 349 rural municipalities. Instruments: See Table 2. Individual level control variables (panel A): age, age squared, gender, an indicator for speaking Swedish as one's mother tongue, six categories for socioeconomic status (entrepreneur, white-collar worker, blue-collar worker, assisting family member, out of labor force) and four categories for the sector of employment (agriculture, manufacturing, construction and services). Municipality-level control variables (all panels): see Table 1. Standard errors in panel A are clustered at the 1939 municipality of residence level.