Where russians should live: a counterfactual alternative to Soviet location policy

Mikhailova, Tatiana

New Economic School

13 January 2012

Where Russians Should Live: A Counterfactual Alternative to Soviet Location Policy*

Tatiana Mikhailova

New Economic School, Moscow

January 13, 2012

(First version: 2003)

Abstract

This paper investigates the extent of distortions in Russia's spatial economy that are inherited from the Soviet system. Using Canada as a benchmark for spatial dynamics of economic activity in a market economy, I construct the spatial allocation of population that would result in Russia, given its initial conditions and existing regional endowments, in the absence of Soviet location policy. The results show that Siberia and the Far East were overpopulated by about 14.5 million people by the end of the Soviet period. Overdevelopment of Siberia comes at the expense of the European area of the country. This discrepancy persists, even after adjusting the simulated counterfactual allocation for WWII.
JEL classifications: R1, P5

*tmikhail@gmail.com. I would like to thank Barry W. Ickes, Eric Bond, Edward Coulson, Regina Smyth, Herman Bierens, Alexei Deviatov, Clifford Gaddy, Susumu Imai, Vijay Krishna, Joris Pinkse, Sergei Izmalkov, workshop participants at Cornell University, Midwest Economics Meeting and Penn State, and anonymous referees for valuable comments. I am grateful to Marjory Winn for help with data. All errors are mine.

1 Introduction

Throughout the 20th century Soviet policies moved population and industry *en masse* from west to east, from the European area of the country to the regions east of the Urals, the cold and remote territories. As a result, the geographical allocation of productive resources in present-day Russia still carries this legacy of Soviet investment decisions.

My goal here is to measure the extent of Soviet distortions on Russian economic geography. Qualitatively, we know that Siberia and the Far East are overpopulated compared to the counterfactual because of the Soviet planning. What we do not know is by how many people these territories are overpopulated. In order to estimate the size of the Soviet distortions, it is necessary to know what would have been an undistorted allocation. To that end, I construct a hypothetical counterfactual allocation of industry and population across geographical space that would have resulted in Russia if it were developing under market conditions.

Why does geography matter? The role of geographical endowment in economic development is a common topic in the economic literature. Cross-country empirical studies find a positive association between growth and proximity to other markets, access to seashore, land quality and mild climate. Gallup, Sachs & Mellinger (1999) conclude that both a hot climate and location away from the seashore hinder economic performance. As Bloom & Sachs (1998) point out, the hot climate of Central Africa is responsible for disease transmission and hinders economic development. Rappaport & Sachs (2001) find a present-day positive productivity premium for U.S. counties located near the seashore and navigable waterways. Gallup, Gaviria & Lora (2003) investigate a variety of geographical factors that contribute to developmental inequality in Latin America. Acemoglu, Johnson & Robinson (2001) propose an explanation for developmental inequality that is based on the quality of institutions, but even in their story the evolution path of institutions is initially determined by physical geography. Redding & Venables (2004) show the importance of access to foreign markets for economic growth, and that market access is a function of geographical location.

Still, physical geography is just an endowment. The fixed characteristics of physical geography are only constraints on economic activity. People, capital, and technologies are mobile, they can adjust to these constraints, and so can mitigate their effect. On the other hand, if geographical endowment is not used efficiently, if these mobile productive resources are misallocated in the geographical space, then economic losses will occur. Spatial inefficiency not only adds to production costs, but when regional comparative advantages are not exploited and unnecessary transportation and communication expenditures are incurred, it also leads to consumption losses. There are extra costs associated with people living in unsuitable places.

Russian geography is unique in two distinct aspects. The first aspect is physical geography (e.g. size, climate, location). Russia's position on the globe can hardly be characterized as favorable. Climatic conditions are harsh. Resources and population are dispersed over a vast territory. The few natural transportation routes (rivers, seas) are located unfavorably for both internal and international trade, leading to the prevalence of costly land transport. Natural resources, although abundant, are primarily located far from population centers in the most severe environments with barely any infrastructure. Russia's size alone is a source of higher transaction costs: transportation and communications must span greater distances. All of these factors drive production costs up, leaving Russia at an absolute disadvantage. The unfortunate geography of Russia and its impact on economic performance have been noted by Lynch (2002) and Parshev (1999).

The second aspect is the extent of Soviet distortions in the spatial economy. Not only does Russia have an unfavorable geographical endowment, but it also uses this endowment badly! With Russia's large size comes a higher cost of errors: if economic activity is misallocated, it is more likely to be *significantly* misallocated and, given Russia's unfavorable geography, such misallocations can be costly.[1]

While Russia's physical geography cannot be changed, the spatial allocation of industry and people within the country can be. Thus, there is room for appropriate policy for

[1]The economy that Russia inherited is probably the most distorted among the Newly Independent States, not only spatially but also structurally. For example, the majority of the defense industry was located in Russia. Gaddy (1996) puts Russia's share of the Soviet Union's defense employment in 1985 at 71.2%, and the share of the population at 51.8%.

mitigating both the extent and the direction of Soviet distortions.

In this paper, I estimate a series of location choice models for population and manufacturing employment based on Canadian data. Then, using these estimated coefficients and starting with data on the Russian Empire in 1910, I forecast the regional shares of population and manufacturing in the USSR in 1990. That is, I construct a counterfactual path of regional population (and manufacturing employment) that would have resulted in the USSR without the distortions of the Soviet planning.

I find that Siberia and the Far East are overpopulated by 14.5 million people (or about 42% of the 1990 population), which comes at the expense of the European part of the country. I estimate that manufacturing employment in the eastern regions should have been at about one third of the actual 1990 level. Also, after adjusting for the impact of WWII, I find that the war cannot explain these differences.

This paper is organized as follows. The next section discusses Russian geography, climate, and the ideology of Soviet location policies. Section 3 describes the general idea behind my analysis. Section 4 describes the empirical setting, and section 5 describes the data. Section 6 outlines methodology and discusses estimation issues. Section 7 presents the results, and section 8 concludes.

2 Russian geography, climate, and history

Russia stands out among other countries not only in its physical geography but even more so in its population geography: the distribution of people over the territory, and the dynamics of that distribution over the 20th century. The peculiar fact of the climatic geography of the Eurasian continent is that winter isotherm lines resemble lines of longitude rather than latitude. Thus, in the process of populating Siberia and even the Urals, people were moving across isotherm lines from warmer to colder places (see Figure 1). On the other hand, summer isotherms look "normal", so migration to Siberia does not reduce people's exposure to summer heat.

Climate and natural resources are not the only factors that are responsible for population geography: people settle where better access to markets (higher market potential) gives them better economic opportunities. People in cold countries – Sweden, Norway, Finland, and Canada – are concentrated along the seashore and in the south, where temperatures are relatively mild and access to foreign markets is easiest. In contrast, Siberia loses in terms of market potential: it is too far from the international trade routes and from the major historical population centers in Russia itself. The density of population is low, and the cities are isolated from one another by distance.

Notwithstanding harsh climate and poor market access, the population of Siberia and the Far East has increased from 5.5% in the Russian Empire in 1910 to about 13% in the USSR in 1989 (25% of the Russian Federation total). Along with population migration came industrialization of these regions. Its share of manufacturing employment increased from 4.8% in the Russian Empire to 12% in the USSR (and 20% within the territory of modern Russian Federation). Until the very end of the Soviet era, Siberia and the Far East were the primary destination for investment. Ozornoy (1991) (Table 2, page 386) shows the regional shares of gross fixed investment in the USSR: the share of the non-European part of the RSFSR from 1976 to 1988 was between 42 and 44.4% of the Soviet total.

One crude yet persuasive piece of direct evidence that the spatial pattern of Soviet development was unusual comes from the dynamics of the temperature per capita (TPC) index, proposed by Gaddy & Ickes (2001), and discussed extensively in Hill & Gaddy (2003). Territorial temperature aggregations are one measure of a country's climatic endowment, and TPC describes how this endowment is used. We can define the TPC of country k as:

$$TPC_k = \sum_j \eta_j \tau_j, \tag{1}$$

where η_j is the share of population and t_j is the mean temperature in region j. TPC typically is measured for a given month – in our case, January, the coldest month of the year. Regions are usually basic administrative units: provinces, *oblasts*, or states. In essence, TPC is a countrywide average temperature aggregated over the spatial distribution of population.

Clearly, a country's TPC is not constant over time. If people migrate from colder to warmer places, TPC rises. Thus, a change in TPC is an index of the climate-related effect of spatial economic evolution. It serves as an aggregate measure of Soviet distortions, because cold is synonymous with remoteness and low density of population in Russia. TPC dynamics shows that Russia was cold at the beginning of the century, and it got even colder through the Soviet years (Figure 2).[2] At the same time, market economies were gradually "warming up".

Of course, the fact that Eastern parts of Russia had been so aggressively developed during Soviet times does not by itself prove that this was not economically efficient. Remote regions should develop even in a market economy if and when technology makes that cost-effective. Direct comparisons with spatial population patterns in other countries – Canada and parts of Northern Europe – are illustrative at best. Russia is different: unique economic and historical factors can potentially provide justification for Siberian development.

Soviet spatial policy: inefficient?

Rodgers (1974) classified ideological principles of Soviet investment policy into three groups: efficient exploitation of resources, equality in geographical pattern of development, and priority of military considerations. In terms of geographical location, the first group dictates a minimization of transportation costs (i.e., locating production near natural resources or consumers). The second group calls for intensified development of the least industrialized areas. The third group gives priority to locations farther from international borders, where production facilities will be safer in the event of war.

All of these ideological principles favor Siberian development. But only the first group takes into account economic efficiency (to the extent that it was possible in the Soviet planned economy, where true economic costs were not observed). The only argument in defense of the Soviet Siberian development might be that it was worth it to locate production closer to the sources of primary materials.

Dienes (1972) shows that even in terms of Soviet prices, average factor productivity on the periphery (including Siberia, the Far East, and Central Asia) was lower than in the

[2]Here Russia refers to the territory in the modern borders of the Russian Federation.

European parts of the USSR. Thus, we can conclude that there has been overinvestment in the eastern regions of the USSR. But we still do not know by how much. Certainly, not all investment in Siberia and the Far East was a waste, given their enormous wealth of natural resources.

History provides other arguments in favor of Siberian development. At the beginning of the 20th century, Siberia was already populated and had several major cities. Even though location fundamentals in Siberia are poor, the path-dependence argument justifies why economic agents might rationally choose to locate there: to exploit already existing agglomeration effects, i.e. to be near other agents. Then, WWII destroyed infrastructure in the western parts of the Soviet Union, and industry shifted to the Urals in the 1940s. This shift was due in part to political decisions of Soviet authorities, but even without political pressure we could expect to see a similar effect in any kind of economy.

Migration trends during the transition period were the opposite of what happened in Soviet times. This again suggests that the Eastern regions were "overdeveloped", but it is neither a conclusive test nor a measure of overdevelopment. There was no theoretical reason for misallocations to reverse themselves after the Soviet period ended, because of the path-dependence of spatial evolution. Agglomeration effects continued to work in Siberian cities. Furthermore, a multitude of factors affect intensity of migration flows in the short-run. For example, a poorly performing region may be locked in a poverty trap: people are credit constrained *and* poor and thus unable to move out (Andrienko & Guriev (2004) proposed this explanation for the low mobility of residents in depressed Russian regions). Although negative net migration can pinpoint the most evident regional problems, it cannot show the degree of Soviet distortions.

Thus, although there is evidence that Siberia and the Far East were overdeveloped, the extent of the Soviet distortions is not clear. Regional endowments and the unique historical circumstances specific to Russia/USSR must be factored in to create an interesting counterfactual.

3 Idea

My idea is to construct a counterfactual that uses Canadian behavior as a benchmark for the spatial dynamics in a market economy, but to apply it to Russian initial conditions and endowment. Using Canadian data, I estimate a model that characterizes the dynamic links between spatial structure of the economy, on the one hand, and initial conditions and regional characteristics, on the other hand. This model is then applied to Russia to produce the counterfactual "market" allocation.

My main concern is the impact of Soviet location policy on the economic geography of the Russian Federation, primarily on Siberia and the Far East. However, for the most of the century Russia had been a part of the common market of the Soviet Union. In order to correctly account for the possibility of interregional migration, the projections must be applied to the territory of what was the Russian Empire, then the Soviet Union, and now the Newly Independent States.[3] Figure 3 shows the administrative borders for the territory of the former Russian Empire and the former USSR (for the list of provinces, see Tables 7 and 8).

I am particularly interested in the joint allocation of population and footloose industry over the territory. Soviet authorities did not just develop mining operations in Siberia, they also built manufacturing facilities there. Clearly, sectors that exploit regional resource endowments are allocated similarly in any kind of economy. Agriculture develops where land is fertile for crops or for livestock. Mining is located at the sites of natural resource deposits. Because manufacturing is *a priori* mobile, to measure missallocations my counterfactual has to account for the factors that drive the regional allocation of manufacturing as well as population.

[3] In different times the term "Russia" referred to different territorial entities. Prior to 1917, Russia was the Russian Empire. During Soviet times, "Russia" was an official short name for the Russian Soviet Federated Socialist Republic – a part of the Soviet Union. After the breakup of the USSR, Russia became a synonym for the Russian Federation - one of the Newly Independent States. The balanced panel dataset that I work with covers all of the regions that were part of both the Soviet Union and the Russian Empire, and goes beyond the borders of the modern Russian Federation.

Why Canada?

Canada is the obvious choice for a benchmark because there is no other country in the world closer to Russia in climate and size.[4] Both economies possess and export abundant natural resources. More importantly for the discussion of spatial population dynamics, both were frontier countries at the beginning of 20th century. Both countries possess vast undeveloped amounts of land. And, throughout history, Russia was expanding eastward and Canada was colonizing its west.

The turn of the 20th century was marked by the construction of transcontinental railroads in both countries at nearly the same time (completed in 1885 in Canada and in 1905 in the Russian Empire). In both countries, the railroad (and a sharp decline in transportation costs that ensued) facilitated a wave of migration to the frontier territories. In Canada from 1901 to 1911, the population of Saskatchewan and Alberta quadrupled, ehile British Columbia's doubled. During the next decade these three provinces grew in population by 50% more (or by nearly half a million people overall). At the same time, the population of the Siberian provinces grew by 50-60%, or by about 2 million people from 1897 to 1910.[5]

At that time, both countries were still predominantly agricultural. The 1901 Census of Canada places 42% of the labor force in agricultural occupations, 17.3% in manufacturing, and 1.6% in mining. According to the 1897 Census the Russian Empire was somewhat less industrialized, with 55% in agricultural occupations, 14% in manufacturing, and 0.6% in mining.[6] The sectoral structure of the economy determined the nature of migration. In both countries railroad construction made vast spaces of unused agricultural land attractive to migrants.

The distinctive feature of the Canadian economy is its close ties to the US, both a

[4] Canada is similar to Russia not only in climate, but also in the share of its territory north of the Arctic Circle.

[5] We have to keep in mind that the Russian frontier was much better populated, even prior to the railroad construction. In 1897 the population of Siberia was about 6.5 million people, practically the size of Canada's.

[6] A statistical summary in the "Yearbook of Russia, 1910" gave an approximate estimate of 75% of population in agriculture, but this figure includes all the family members of agricultural workers. Canadian censuses traditionally did not record the dependents of farmers as gainfully employed. Therefore, to make a valid comparison between two countries I cite the occupation shares in the Russian labor force, i.e. only among the people who listed a primary occupation in the 1897 census. Only 55% of them were agricultural workers.

friendly neighbor and a large market. In general Canada is in a better position for both internal and international trade than Russia because of its natural water transportation routes and good access to oceans. But the Russian Empire was also in a position to benefit from a large neighboring market: Europe. Of course, the Russian Empire (and, later the Soviet Union) shared borders with an extremely diverse set of countries, and relationships with European neighbors were uneasy at times. At the beginning of the century the major trading partners of the Russian Empire were Germany and United Kingdom. On the other hand, the prevalence of land transport in Russia makes shipping distances very costly. It is not clear whether being closer to the western border in Russia yields a stronger or weaker trade advantage than being close to the U.S. border in Canada.

One marked difference between the two countries is in GDP per capita. At the turn of the 20th century, an average Canadian was 2 times richer than an average Russian.[7] This difference is crucial to the extent that income affects the mobility of capital and labor. One might argue that because of higher income, the Canadian population was better equipped to take advantage of migration opportunities. Indeed, population mobility in Canada today is higher than in the Russian Federation today, and it was somewhat higher (in proportional terms, but not in total numbers) at the beginning of 20th century. Still, endowing "counterfactual Russians" with high "Canadian mobility" provides more, not less, justification for the migration from the traditional core to the periphery of the country. If anything, I risk underestimating Soviet distortions.

There are two marked differences in the factors that drove spatial evolution of the economy in Canada versus Russia. WWII affected the spatial structure of the Soviet economy enormously, but there is no precedent for such an impact in Canadian history. Another issue is regional diversity. The regions of the Russian Empire (or the Soviet Union) are more diverse than Canadian regions with respect to ethnic composition, human capital, and culture. Cultural differences, most notably traditional differences in fertility rates, had a greater effect on regional shares of the population in the USSR than in Canada. However, I do not believe that this is crucial for my results, because the regions I am primarily interested in

[7] Amazingly, this proportion is exactly the same for 2009 World Bank PPP estimates!

– Russian Federation and especially Siberia – are more homogeneous than the USSR as a whole.

In the following sections, I estimate an empirical model of spatial dynamics of population and industry on Canadian regional data. The equations link regional population growth to past population, industry, and various location characteristics. Then, the equations bearing the fitted values of the coefficients are applied to the Russian Empire's regional data on the initial (beginning of the century, before the October Revolution) population and industry and the same set of regional characteristics. The result of the projections is the counterfactual allocation – specific to the Russian starting point and geographical characteristics, but obtained using the dynamic relationship fitted on a market economy. This procedure is described in greater detail in Section 6. In section 7.4, I correct these projections for the impact of WWII and fertility differences.

Thus, I am assuming not that the spatial structures of different market economies should be similar, but rather that the dynamic forces that affect location choices should be similar. In other words, I do not simply compare the existing spatial allocations in Russia and Canada, but rather look at the changes in structure over time: initial conditions matter.

4 The general framework

There are two basic approaches to modeling location choice. As the rule, the location choices of firms are modeled as a decision to enter one of several possible markets: a firm does not have a location *a priori*. The location choices of people (termed migration) are modeled as a decision to change ones's existing location: the choices of agents and the outcome together depend on the initial regional population structure. I discuss both approaches in greater detail below .

4.1 Location of industry

Carlton (1983) was the first to apply McFadden (1974)'s multinomial logit framework to the industrial location choice problem. Since then, this framework has been widely used for the estimation of regional industrial structure models.

Consider a specific industry consisting of many small firms that are deciding where to locate among a finite number of regions. Let π_{jl} be the profit of firm j in region l. Firms maximize profits, so if firm j locates in region r, then $\forall l$, $\pi_{jr} \geq \pi_{jl}$. Assume that a firm's profit function in a region is a combination of the region-specific component common to all firms and the random shock that is specific to the firm and region. Formally:

$$\pi_{jl} = \tilde{\pi}_l + \epsilon_{jl}. \tag{2}$$

Common component $\tilde{\pi}_l$ encompasses all of the characteristics of region l that make it attractive to firms on average. Random variable ϵ_{jl} captures the attractiveness of region l that is specific to firm j. In essence, it represents the "quality of the match" between j and l: *ceteris paribus* a positive shock raises j's profit from being in region l; a negative shock decreases it.

If we further assume that ϵ_{jl} follows Gumbel law, with cdf $F(\epsilon) = \exp(\exp(-\epsilon))$, then the probability that firm j chooses location r is given by the logit function:

$$P_{jr} = \frac{\exp(\tilde{\pi}_r)}{\sum_l \exp(\tilde{\pi}_l)}. \tag{3}$$

Let Ind_r be the total number of firms in region r and Ind be the total number of firms in the industry, $Ind = \sum_r Ind_r$. As the total number of firms in the industry increases, regional shares $S_r^{ind} = \frac{Ind_r}{Ind}$ converge to probabilities (3). Thus, by taking logs and differencing, we obtain:

$$\ln Ind_r - \ln Ind_l = \ln S_r - \ln S_l = \tilde{\pi}_r - \tilde{\pi}_l. \tag{4}$$

To obtain an equation to estimate, one needs to specify the average regional profit function $\tilde{\pi}_r$, but the profit function is dictated by the question under being considered. To construct industrial concentration indices, Ellison & Glaeser (1997) consider a simple two-component additive form: one component is a composite of all natural advantages of location, the other reflects spillovers that arise from co-locating with other firms. The majority of empirical studies of the determinants of FDI or domestic industrial location use linear hedonic specifications. (For a recent survey, see Arauzo-Carod, Liviano-Solis & Manjón-Antolín (2010).)

Head & Mayer (2004) derive the regional profit function from a theoretical New Economic Geography model. They show that up to a monotonic transformation, the profit function is linear in real market potential and marginal cost. These variables in turn depend on the location of the region, regional incomes and factor prices, and total factor productivity.

Thus I assume that the profit function is linear in the observable regional characteristics and unobservable component:

$$\tilde{\pi}_r = \sum_{k=1}^{n} \alpha_t^k x_r^k + \delta_r. \tag{5}$$

Then,

$$\ln Ind_r - \ln Ind_l = \sum_{k=1}^{n} \alpha^k x_r^k - \sum_{k=1}^{n} \alpha^k x_l^k + \delta_r - \delta_l. \tag{6}$$

Assume that δ_r are zero-mean normal random variables. Then $\xi_r = \delta_r - \delta_l$ are also normal with zero mean. Because regional shares must sum to one, only $n-1$ equations (4) can be written for n regions (one degree of freedom is lost). Choosing region l as a common numeraire and collecting common terms into a constant, the equation to estimate is:

$$\ln Ind_r = \alpha^0 + \sum_{k=1}^{n} \alpha^k x_r^k + \xi_r. \tag{7}$$

If firms are small and *a priori* identical, then regional levels of industrial employment are approximately proportional to regional firm counts Ind_r. In this paper, I consider the location of manufacturing and measure Ind_r by manufacturing employment in the region.

4.2 Population migration

In the literature population migration is traditionally modeled at the level of bilateral inter-regional flows. Person j who lives in region h maximizes utility u over a set of regions and decides to migrate to region r if $\forall l$, $u_{jhr} \geq u_{jhl}$. The set of possible choices l includes the home region (the option to stay); regional utility u_{jhl} includes the cost of moving from h to

l (zero for u_{jhh}). Thus, the decision to migrate depends on the characteristics of both the source and destination regions, and choice sets are different for residents of different regions.

As for industrial location choice, assume that individual shocks are additive and Gumbel-distributed:

$$u_{jhl} = \tilde{u}_{hl} + \epsilon_{jhl}. \tag{8}$$

The share of the population of region h who migrate to r is:

$$M_{hr} = \frac{\exp(\tilde{u}_{hr})}{\sum_l \exp(\tilde{u}_{hl})}. \tag{9}$$

The population of region h is the sum of those who stayed and those who came from the other regions, and it depends on the initial population distribution (indexed by 0).

$$Pop_h = Pop_h^0 \frac{\exp(\tilde{u}_{hh})}{\sum_r \exp(\tilde{u}_{hr})} + \sum_{l \neq h} Pop_l^0 \frac{\exp(\tilde{u}_{lh})}{\sum_r \exp(\tilde{u}_{lr})}. \tag{10}$$

Thus the regional share of population depends on the initial shares and characteristics of all regions. Interregional migration flows have a simple analytical form, but aggregate shares of population do not. The general model is intractable.

Assuming away the cost of moving simplifies the model. If mobility is costless, then $\forall l, r$, $\tilde{u}_{lh} = \tilde{u}_{rh} = \tilde{u}_h$, and equation (10) reduces to the multinomial logit expression:

$$Pop_h = \sum_l Pop_l^0 \frac{\exp(\tilde{u}_h)}{\sum_r \exp(\tilde{u}_r)} = Pop \frac{\exp(\tilde{u}_h)}{\sum_r \exp(\tilde{u}_r)} \tag{11}$$

As in the industry case, under an assumption that regional utility is a linear function of regional characteristics, equation (11) is transformed into:

$$\ln Pop_r = \beta^0 + \sum_{k=1}^{n} \beta^k x_r^k + \eta_r. \tag{12}$$

4.3 Explanatory variables

As a general rule, with forecast models the set of explanatory variables must be small for the forecast to be meaningful and robust. On the other hand, the set must be rich enough to capture the basic trends of regional growth through the 20th century in Canada. There are four main considerations in choosing this set.

First, the set of regional factors must be restricted to those available for as early as 1910 for the Russian Empire, for all of the Soviet period, and for all of the decades from 1911 until 1991 for Canada. The data also must be comparable between two countries.

Second, to capture history-dependence of location choices, I include past levels of dependent variables. Choices of people and firms are interdependent: firms locate where people are, and people migrate toward jobs. Accordingly, past levels of population and manufacturing employment enter the set of right-hand-side variables in each equation. Thus, the estimated equations are linked recursively.

Third, I use only exogenous regional characteristics. Many of the factors that are clearly important for migration (regional wage and income differentials, housing prices, unemployment) or for industrial location (intra-and inter-industry spillovers) themselves depend on the regional allocation of population and industry. Note that we cannot observe what any of these endogenous variables would have been in the "counterfactual USSR." Thus, to be used in the forecast, they would have to be predicted first, which would make the model overly complicated.[8]

Finally, the variables must reflect the main factors of interest. Was exploration of natural resources in remote regions followed by growth of population and manufacturing? Did population or manufacturing employment grow faster when in close proximity to international borders and trade routes? How strong is persistence in regional population shares and shares of mobile industry? Does industrialization (growth of the manufacturing sector) drive pop-

[8] A similar issue arises with respect to man-made infrastructure and natural resource exploration. Both are obviously important for location decisions and endogenous to population. I use 10-year lagged values for all of the variables that describe man-made infrastructure and development of natural resources, which changed throughout the course of the 20th century. For the counterfactual forecast, I take Soviet development decisions as given and use them as inputs for the forecast.

ulation into the region? To answer these questions, the set of x_r has to include information on natural resources, geographical location, past levels of population, and manufacturing employment.

As shown in the work of Head & Mayer (2004) on industrial location, Crozet (2004) on migration, and H. Hanson (2005) on regional real wages, real market potential (RMP) is an important factor that attracts firms and population. RMP is a theory-based composite measure of how large the demand for local goods is, and it takes into account trade costs, average income, and prices in all other markets that can be reached from a given location. Including the precise theoretical formulation of RMP into the set of x_r would make model specification too complicated for the forecast, but factors that proxy for RMP must be included. Past levels of population and manufacturing proxy for the size of regional market; geographical variables (for example, closeness to the international border or port) and infrastructure characteristics (such as railroads) proxy for low trade costs (good market access).

Models of population migration normally include labor market factors (real income or wages, unemployment), and quality-of-life indicators (housing prices, climate, amenities). For the purposes of a forecast, endogenous variables (wages, prices, unemployment) must be expressed as functions of the observed exogenous regional characteristics. Variables that reflect natural advantages in terms of quality-of-life (mild climate, waterways) would tend to lower the local real wage, while variables that proxy for high RMP would tend to increase wages.

The resulting set of explanatory variables includes: past levels of population and manufacturing employment as well as a vector of the regional characteristics of climate and geography, natural resources and land endowment, and man-made transport infrastructure.

Estimated equations

The resulting estimation equations link regional levels of population and manufacturing employment with their lagged values and a variety of regional characteristics:

$$\ln Ind_{rt} = \alpha_t^0 + \alpha_t^{Pop} Pop_{r,t-1} + \alpha_t^{Ind} Ind_{r,t-1} + \sum_{k=1}^{n} \alpha_t^k x_{rt}^k + \xi_{rt}, \quad (13)$$

$$\ln Pop_{rt} = \beta_t^0 + \beta_t^{Pop} Pop_{r,t-1} + \beta_t^{Ind} Ind_{r,t-1} + \sum_{k=1}^{n} \beta_t^k x_{rt}^k + \eta_{rt}. \quad (14)$$

Variable t indexes time. There are 8 time periods; length of the time period is 10 years, the lag of the census data at the beginning of the 20th century in Canada. Year 1911 is a starting point ($t = 0$). Thus, 1921 manufacturing and population are expressed as functions of 1911 data, 1931 as functions of 1921, etc up to 1991. In general, I allow parameters α and β to be different in different time periods, reflecting the fact that the nature of spatial dynamics can change through time.

Estimated parameters $\hat{\alpha}_t^k$, $\hat{\beta}_t^k$ together with data on the initial regional levels of population and manufacturing in the 1910 Russian Empire and the characteristics of Russian/Soviet regions, are then used to construct a forecast of regional shares of population and industry in the counterfactual USSR around 1990.

Section 5 describes the data, and Section 6 discusses the empirical strategy in more detail.

5 Data

For Canada the dataset is a panel with 9 time points: one population census year per decade from 1911 to 1991. For any given year, the data on regional population, manufacturing employment, and other regional characteristics are compiled (see Table 1 for details). These same data are collected for the Russian Empire for 1910 — the starting cross-section for the counterfactual forecast. For some explanatory variables the data also are collected for the USSR for each decade until 1990, and they serve as inputs for the forecast. The final Soviet population (1989) and industrial (1988) censuses are then used for comparing the actual allocation with the counterfactual one that is obtained.

5.1 Spatial units of analysis

The choice of the geographical unit for the analysis is dictated by availability of data. The smallest unit of geographical aggregation for the manufacturing data that is available for the Russian Empire is "*gubernia*" or "*oblast*" ("province" is a most suitable name equivalent), a basic administrative unit. There were 98 administrative units in the Russian Empire in 1910. But only those territories that later belonged to the Soviet Union are included in this sample. Eighteen provinces that were in Poland and Finland, and Karsskaya province that is now a part of Turkey, are excluded. Separate data for Kamchatka oblast and Zakatalskii okrug do not exist, so they are included in Primorskaya and Tiflisskaya provinces respectively. The result is 79 regions. Figure 3 shows a map of the administrative divisions in the Russian Empire, together with the borders of the USSR and the present day borders of the Russian Federation.

One crucial step of this analysis is constructing comparable spatial units (regions) for Canada, because Canadian provinces are too big when compared to Russian ones. I took the data reported at the Canadian census district level and combined census districts into artificial geographical units similar to Russian ones. The main difficulty here stems from the fact that census districts were changing over time as the Canadian territory was gradually developed and populated. The census of 1911 had fewer and larger by area districts. Wherever large 1911 districts were later divided up, I had to combine districts for the later years in order to maintain the panel. Most of the sparsely populated territories – northern parts of Prairie Provinces, Yukon, North-West Territories, mountains of British Columbia – were single-district territories in 1911.

Districts located in the densely populated Canadian South were also sometimes revised. On several occasions, the new and old districts overlapped significantly. In these cases, I typically merged overlapping districts. The exception was several districts in Northern Quebec that were revised several times throughout the century. The boundary changes were significant there, but the overlaps covered territory that had extremely low population density. In these cases, I counted a new district in the same geographical unit where the largest city of the district had been located. I do not believe that significant error was introduced in the process, because the overlaps covered only small distant villages.

Next, I merged the geographical units to construct a set of regions that would closely resemble Russian imperial administrative units in both size and spatial pattern. Provinces in Russia are mainly formed around an urban center. In Eastern Siberia and the Far East, administrative borders follow the topography of the terrain. Provinces in the European part of the Russian Empire were smaller and had higher population density. In Central Asia, Siberia, and the Far East, the provinces are large and sparsely populated. Where possible, I applied similar principles to Canadian districts.

Generally, I combined small districts (counties) in southern Ontario, southern Quebec, Nova Scotia, and New Brunswick so that, where possible, each division covered a large or mid-size city and the counties around it. Districts in the Prairie Provinces often are rectangular, and they can be divided naturally into southern, middle, and northern parts, with a large or mid-size city in each.[9]

The result was 38 regions constructed from 227 census districts in 1911 Canada.

5.2 Population and manufacturing employment

The population data for pre-revolutionary Russia come from the "Yearbook of Russia, 1910." For manufacturing employment data, I used the reports of the 1908 census of industry which give the number of people employed in manufacturing enterprises with 5 or more employees.[10]

The data for population and manufacturing employment for the pre-transition Soviet Union were taken from the Population Census of 1989, and the Census of Industry, 1988. Both employment and population data are available at the level of basic geographical aggregation (by *raion*). The *raion*-level data were regrouped to the boundaries of the pre-revolutionary provinces.

[9] Where districts can be combined objectively in different ways, I tried each of the combinations. The estimation results are not sensitive to the particular choices.

[10] Another source of data on manufacturing employment in the Russian Empire is the report on peoples' occupations by province, published in "The Yearbook of Russia, 1910." The number of people reported to be in manufacturing, mining and crafts is somewhat higher than counted by the census of industry, because self-reported occupations include people in small establishments. The results are robust to the choice of data source; the difference is a fraction of a percentage point.

The main sources of data for population and manufacturing employment are the Censuses of Canada for each decade from 1911 to 1991. For 1921-1931, census publications give industry data only for cities with populations greater than 15,000, but not for census districts. These years were dropped from the sample for the industry equation and were replaced by 1911 data in the population equations for 1931 and 1941. Data for the year 1981 are not included in census publications, but are available from “Manufacturing Industries of Canada: Geographical Distribution,” Regional and Small Business Statistics Section, Manufacturing and Primary Industries Division, Statistics Canada, 1982. Unfortunately, due to confidentiality issues, several small mono-industrial census districts are not listed. Thus, the quality of data for 1981 is substantially lower.

5.3 Regional characteristics

The regional characteristics used for the estimations (for Canadian regions) and for forecasting (for Russian regions) are summarized in Table 1.

Area , temperature, agricultural land quality, having a port, and distance to the prime city all represent the inherent characteristics of a region: size, climate, quality of soils, accessibility, and remoteness. If the region is not landlocked, then the largest city is usually a port, so the “Port” variable also can be considered inherent to geographical location. These variables do not change from period to period.

Railroads and trade route are characteristics of not only the location (accessibility) of a region, but also the level of infrastructure development. Because the structure of a railroad network is changing over time and is endogenous to the population distribution, it is imperative to use the lagged number of railroads in the Canadian regressions. I also use the actual dynamics of Soviet railroad network development as an input for the forecast, because it is obviously impossible to know where railroads would be built in a counterfactual world. By doing so, I am implicitly assuming that Soviet infrastructure investments were rational from the market economy point of view (even though there are obvious reasons to suspect that they were not). This may introduce bias “in favor” of the Soviet development pattern, but the results show that Siberia is overpopulated nonetheless.

Variable	Description
Area	Land area, sq. km.
Temperature	Average January temperature in the largest city in a region, °C.
Distance to the prime city	Direct (straight line) distance from the largest city in a region to Toronto or Moscow.
Railroads	Number of railroad branches leading from the largest city in a region.
Coal mining	1, if a significant amount of coal was mined.[a]
Metals mining	1, if a significant amount of any metal ores were mined.
Oil extraction	1, if a significant amount of oil was extracted.
Timber cutting	1, if at least 1/3 of the territory is covered with forest and a significant amount of logging was taking place.
Port	1, if the largest city is a port. Included are Canadian ports on the Atlantic and Pacific oceans, on Great lakes, and on St. Laurence river, and Russian ports on all seas and the Caspian lake.
Trade route	1, if there is a direct (not through some other region) transportation route (road, railroad, or waterway) abroad.
Agricultural land	1, if at least 1/3 of the land is classified as "having no major obstacles for agriculture." For Canada, corresponds to land type A and B in the agricultural lands classification.
Urbanization	% of urban population in a region in 1911

[a] For natural resources (coal, metals, oil, and timber) "significant" means that the region is a net exporter of the resource.

Table 1: Regional Characteristics

Presence of a trade route is treated as exogenous, since all the routes connecting Canada and US are traditional transport routes and were formed prior to the beginning of the century. The same is true for the USSR and its transport connections with the neighbors.

Natural resources endowments are characterized by four variables: presence of coal mining operations; presence of oil extracting operations; presence of any metal mining operations; and presence of timber resources. All of these are dummy variables indicating only the presence of active resource-extracting operations in a region. To track changes in mining operations and to avoid possible endogeneity, the dummy variables vary with time, and the lagged values are used. The sources of data for Canada are statistical yearbooks (inter-provincial trade) and natural resource maps. For the USSR, they are natural resource maps and various Soviet public historical news records.

The choice of dummies is a compromise. The truly relevant factors for the population or industry growth in a region are the potential economic profits that could be obtained either from natural resources or land, or the value of positive externalities provided by the existing infrastructure or favorable location. Characterization of these factors by the use of dummy variables is certainly a simplification of reality, but unfortunately it is necessary. I had to choose regional characteristics bearing in mind that data for both Canada and the Russian Empire (or the USSR) had to be collected. Using more informative measures for a region's natural resource endowment (the amount of extractable resources, for example), its land value and infrastructure was possible for Canada, but comparable information on Russian/Soviet regions does not exist in the public domain,[11] or at all. Only simple dummy variables can be constructed using open sources, including maps and statistical publications.

6 Empirical strategy

This section describes the selection of the best forecast model used with the Canadian data, the estimation and the forecast procedure in detail. Any forecast problem involves model selection: that is, choosing a subset from the set of potentially relevant explanatory variables.

[11] For example, estimates of extractable natural resources were USSR state secrets.

Because I have a series of recursive forecast equations, the choice of the correct model is crucial. The next subsection describes the estimation procedure for any given model specification.

6.1 The procedure

Step 1 Estimate the selected dynamic panel model for Canada to obtain the fitted values of the model parameters.

I estimate the system of equations (13) and (14) on the Canadian panel data. For a specific selected model, restrictions on the parameters are imposed accordingly. Examples of such restrictions are: $\alpha_t^k = 0$ if factor k is excluded from the equation for period t, or $\forall t$, $\alpha_t^k = \alpha^k$ if the coefficient before factor k is restricted to be constant over time.

Because data on regional manufacturing in Canada are missing for 1921 and 1931, they are replaced by 1911 data in the population equations for 1931 and 1941 and in the manufacturing equation for 1941. Alltogether I have eight equations to estimate for population and six for manufacturing. If all parameters are decade-specific, the dynamic panel model breaks down into period-by-period equations. Nevertheless, they are estimated as one system because equations (13) and (14) represent seemingly unrelated regressions. The error terms for industry and population in the same region and the same time are likely to be correlated: a positive shock to the region's population is likely to coincide with a similar shock to manufacturing employment.

Step 2 Project the estimated relationship onto Russian/Soviet data.

The forecast is recursive. Beginning with the actual data for population and industry in 1910, I use the population equation for 1921 to calculate the projected population in 1921. Then, with the projected 1921 population and the 1910 actual industry data I obtain the projected population for 1931, and repeat the procedure for 1941. Industry in the year 1941 is a function of 1931 projected population and 1910 actual industry data. For each decade from 1951 until 1991, projected population and industry are computed using the estimated coefficients and the projected population and industry from the previous decade. The first two projection equations for the population and industry (regional index is omitted, F indexes forecasted values) are:

$$\ln Pop_{1921}^{RusF} = \beta_{1921}^{Rus} + \hat{\beta}_{1921}^{pop} \ln Pop_{1911} + \hat{\beta}_{1921}^{ind} \ln Ind_{1911} + \sum_k \hat{\beta}_{1921}^{k} x_{1911}^{k}, \tag{15}$$

$$\ln Ind_{1941}^{RusF} = \alpha_{1941}^{Rus} + \hat{\alpha}_{1941}^{pop} \ln Pop_{1931} + \hat{\alpha}_{1941}^{ind} \ln Ind_{1911} + \sum_k \hat{\alpha}_{1921}^{k} x_{1911}^{k}. \tag{16}$$

This procedure is repeated until the year 1991.

Note that the constant terms β_t^{Rus} and α_t^{Rus} are not equal to the intercepts in the equations (13) and (14), respectively, estimated on the Canadian data. In multinomial logit models the shares of different regions must sum to one, so one degree of freedom is lost to that additional condition. When projecting the model onto Russia, and properly calculating the relative shares of all the provinces, we need to account for that loss. Therefore, the values of β_t^{Rus} and α_t^{Rus} can be obtained from the condition: $\sum_{l=1}^{79} POP_{l,t}^{RusProj} = POP_t^{RusActual}$, where $POP_t^{RusActual}$ is actual total population of the Soviet Union in year t.[12]

Next, I use the projected 1921 values to obtain projections for 1931, and so on until 1991. The 1991 results present an alternative spatial allocation for Russia that would have occurred if its development followed the Canadian path. This is the counterfactual allocation I sought: it is free of the shocks and disadvantages that are specific to Soviet history.

This procedure also can be applied to the Canadian data for 1911, to obtain projected values for 1991 and to compare them with the reality, using a chosen criterion. This is one way to evaluate the quality of the model. I discuss other model selection issues below.

6.2 Model selection

Different choices of explanatory variables and other restrictions on the coefficients in estimated equations(13) and (14) lead to different forecasted counterfactuals. As with single-equation models, it seems that the specification that provides the best fit for each individual

[12]It is not necessary to obtain the value of the constant term for each year. Since the relationship between past and present population (or industry) is linear in logs, the constant added to all the observations does not change the relative "weight" of different regions. Only the terminal-year constant term is of interest.

period does not necessarily provide the best overall forecast.

Forecast model is commonly selected by the quality of out-of-sample fit or by information criteria. Out-of-sample forecasts are not feasible here, because there is no alternative, parallel-world Canada on which to test the performance of a forecast model. Nor is it possible to split the sample of Canadian regions into a fictitious "in-sample" and "out-of-sample" either: the sample is small, the regions are diverse, and any single sub-sample would not be representative.

Information criteria can be used to resolve the trade-off between too many and too few explanatory variables in small samples: additional variables improve goodness-of-fit but also may "over-explain" realized shocks, resulting in biased coefficients and poor predictive qualities. In the case discussed here, the forecast is recursive, and it is not clear whether the model selected by the information criteria works well overall. This is because the forecast errors accumulate.

When I use the predicted values of population and industry as inputs for the next-period forecast, the errors for the same region in different periods can either accumulate or cancel each other out. Indeed, a common issue in cross-country or regional panel models is autocorrelation in the residuals in the time dimension.[13] Certain persistent factors that affect location and are not included in the set of explanatory variables can generate positive autocorrelation of the residuals in the Canada panel. And, if the residuals are positively correlated, then the recursive forecast accumulates and inflates the errors that have occurred over time.

One way to deal with persistent omitted variables in structural panel models is with individual (fixed or random) effects that allow us to estimate parameters correctly. In this analysis, the individual effects are useless: they cannot be used for the forecast, because they are not observed for the USSR. Also, I am not interested in unbiased estimation of structural parameters; I am interested in projections. The forecast, which is a linear projection, is unbiased under the central assumption that both observable and unobservable factors would have affected location decisions in the same way in the "counterfactual USSR" and in Canada.

[13]See Islam (1995) for the discussion of the common estimation issues in panel models of growth.

I can instead use in-sample forecast criteria for model evaluation. Based on the estimated model, I predict regional population in Canada in 1991 via the recursive procedure from the 1911 population, manufacturing, and regional characteristics. I then compare this 1991 estimate with the actual data using a chosen criterion. For example, for model selection I might use (A indexes actual values, F - forcasted):

- Minimum sum of relative squared errors (relative SSE): $\sum_r \left(\frac{POP^A_{r,1991} - POP^A_{r,1991}}{POP^A_{r,1991}} \right)^2$.
- Minimum sum of absolute squared errors (absolute SSE): $\sum_r \left(POP^A_{r,1991} - POP^F_{r,1991} \right)^2$.

I select the model (chosen set of variables) that gives the best result according to one of these criteria.

I chose "relative SSE" as the criterion for the main model, because it seems to yield best results in terms of the spatial dynamics in Canada. Because the main focus of this paper is the population of the Russian east, I must make sure that the model predicts population growth in peripheral regions properly. The model that minimizes relative SSE is selected to fit regions irrespective of their population with equal weights. Absolute SSE yields a model that fits the populous metropolitan areas better. For my purposes, the relative SSE criterion is a better choice.

Ideally, the search for the best model would involve estimating the projected Canadian population for all possible models, and then selecting the model that minimizes the selection criterion. However, with 14 explanatory variables for each period and for each dependent variable, the number of possible combinations is astronomical – $2^{14*(8+6)} = 2^{196}$ possible models. Instead of exhaustively searching through all specifications, I use an iterative algorithm to find the best model. The algorithm sequentially eliminates or adds explanatory variables reducing the SSE in each step. A detailed description of the algorithm is in Appendix A.

The drawback of the in-sample selection method is similar to the critique of data mining procedures. I am fitting Canadian dynamics "too closely," picking up events that might be random noise rather than features of spatial evolution. The Russian counterfactual thus in a sense is "too Canadian." But this is the nature and the main assumption of the counterfactual exercise, because there is no way to distinguish between random and systematic events by observing just one realization of Canadian history.

Clearly, the minimum relative SSE model (and the minimum absolute SEE model) prevents us from over-accumulating errors in the recursive forecast. Intuitively, the model selects a set of explanatory variables that are strongly correlated with omitted unobservables. In turn, we must keep in mind that the estimated coefficients of the model are biased away from its true structural parameters. They describe correlations, but they do not imply any structural interpretation.

6.3 Robustness

To ensure that the results I obtain are not driven solely by the specific model I select, I estimate several different models, make projections on Russian/Soviet data, and compare the results. Overall, my main message – that Siberia and the Far East are overpopulated and overindustrialized – remains true for all the models. In addition to the main model (chosen by relative SSE criterion, indexed by (1)) I look at the following models:

(2) A model, that minimizes the sum of the absolute squared errors (absolute SSE)

The next models, (3)-(8) are chosen without in-sample forecast evaluation, and they range from simple to more complex.

(3) Naive extrapolation of the 1911 regional shares of population and industry onto 1991.

(4) All regional characteristics are included, but model parameters are fixed over time.

(5) Parsimonious model: only past levels of population and industry are included in the regressions.

(6) Only past levels of population and geographical characteristics are included (no natural resources or railroads).

(7) Complete model: all of the variables are included.

(8) Model chosen by Akaike criterion: I reduce the complete model by consecutively excluding variables until the Akaike criterion is minimized.

Model (3) is the best guess about a counterfactual Russia if information about the spatial evolution of Canada in the 20th century is not used at all. Model (4) predicts Russian/Soviet population recursively based on Canadian dynamics, but is very restrictive. It implies that a given factor affects the location of people or manufacturing in the same way throughout the 20th century. Obviously, there is no theoretical reason to expect this.

For the rest of the models I relax this assumption. I calculate the results for models (5)-(7) to demonstrate that there is no particular group of variables that drives my main result. Finally, model (8) is chosen so as to minimize information criterion.

Another essential feature of a good forecast model is the absence of systematic spatial bias. If the differences between projected and actual population values do not appear to be spatially random, then the model is likely biased and should not be used. I calculate Moran's I statistics (with contiguity spatial weight matrix) for the forecast errors for Canada to check for spatial autocorrelation. My results should also be examined for any apparent geographical biases.

The next section presents in detail the results for the main model, compares the results obtained via other models, and discusses their robustness.

7 Results

7.1 Projection models: performance in Canada

The summary results for all eight models are presented in Table 3. The top part of the table summarizes the results of the projections for Canada (in-sample).

Rows a) and b) show the R^2 measures for Canadian projections:

$$\text{Absolute} \quad R^2 = 1 - \frac{\sum_l (Pop^A_{l,1991} - Pop^F_{l,1991})^2}{\sum_l (Pop^A_{l,1991})^2},$$

$$\text{Relative} \quad R^2 = 1 - \frac{\sum_l \left(\frac{Pop^A_{l,1991} - Pop^F_{l,1991}}{Pop^A_{l,1991}} \right)^2}{N}.$$

Canadian population (absolute levels) is predicted reasonably well, even by the "naive" extrapolations of 1911 (model (3)). The model places 86% of Canadian population where it should be, which is not surprising, because major population centers in Canada (except for western cities, such as Vancouver, Calgary, Edmonton) formed and grew prior to the beginning of 20th century, and have retained their significance up to the present day. Canada's present population distribution is predicted well by its past distribution. Of course, more sophisticated models fit the reality even more closely.

Practically all of the models predict Canada as "colder" than it actually is: predicted TPC (row c)) is consistently below actual TPC. This suggests that in the Russian projections the systematic error likely would be on the side of "colder" population allocation; i.e., it would put more people into Siberia than should be there. Thus, if anything, my results underestimate the degree of Soviet distortions. Moran's test (row d)) cannot conclusively find spatial autocorrelation in any but the most primitive models.

Figure 4 shows projection errors for the main model (1). Positive and negative errors are distributed fairly evenly across the territory. There is no immediately visible bias either for or against any particular part of the country.

The model is worse at predicting population in large metropolitan areas. Several regions with major cities (Vancouver, Toronto, Montreal, Edmonton) have been growing rapidly during the 20th century and have a strongly positive unexplained error: their actual population is higher than projected. The other large cities (Winnipeg, Halifax, Ottawa, etc) have either negative or near-zero errors. On average, the population of the largest metropolitan areas is underestimated. But that should be expected from any model: locations with a positive unexplained shock to growth have become leading metropolitan areas. Also, even if systematic bias against large cities exists, it works against predicting more population in the western part of Russia and toward predicting more population in the frontier east. Again, this tends toward underestimation of the extent of Soviet distortions.

7.2 Estimated Canadian dynamics: discussion

Tables 4 and 5 present the estimation results for the main model. What facts about Canadian spatial dynamics can we take from the estimated coefficients? In particular, does the common rhetoric about Soviet Siberian development fit the Canadian data? Do resource-rich regions indeed provide easy access to primary materials, and hence attract manufacturing and population?

Population equations

Given that the parameters are not structural, and that the number of parameters is quite large for the sample size to provide valid inferences, it is still possible to make several observations on Canadian spatial dynamics. First, the most important factor in population growth is past level of population. This is not surprising: the spatial structure of the economy is history-dependent and very inert. The urbanization rate in 1911 has a consistently positive coefficient, suggesting that areas that were settled and urbanized prior to the beginning of the century continued to attract people at a higher than average rate. Second, the coefficient of land area is mainly positive, *ceteris paribus* indicating a diffusion process: people tend to spread across the territory over time, rather than concentrating in large agglomeration points. This pattern is quite plausible for the countries going through territorial expansion or, as in the cases of Russia and Canada, settling sparsely populated territories. Third, resource-rich regions do not accumulate population faster than the average region. Finally, variables that proxy for trade possibilities – communications and the market potential (number of railroads, route abroad and ports), which are presumably advantageous characteristics – by themselves are not correlated with more rapid population growth. Most likely, the positive influence of these factors is already built into past population levels.

Lagged industrial employment is also positive and significant for the middle part of the century: i.e., population grew faster in industrialized regions. However, it is negative for 1921, and close to zero for 1931, 1941, and 1991. Poor quality of the 1981 employment data may explain the zero coefficient in 1991. The 1931 and 1941 equations use 1911 data for past industry in place of the missing 1921 and 1931, so the variation in coefficients is expected. An alternative explanation is that the 1911-to-1941 time period also covers two major historical events: the massive migration to the Western provinces and the Great Plains at the beginning of the century and The Great Depression of the 1930s. Thus, the near zero or negative coefficients for lagged industrial employment and temperature might reflect population movement away from industrial centers and to the regions with colder climates.[14] Interestingly enough, in many time periods temperature itself is not a significant explanatory variable. Movement of population towards warmer areas is explained well enough by other factors, for example lagged population.

Industry equations

For industry, the role of lagged variables is reversed: lagged industry is now a significant factor. Unlike population, area does not always have a strongly positive coefficient, and lagged industry coefficients for different years can be either under or over one. Thus, it is not clear whether industrial employment follows the same diffusion-type spatial dynamics as population. Industry growth is correlated with the presence of infrastructure and better market access: the number of railroads has a positive and significant coefficient in two time periods. However, the natural resources variables have predominantly negative or near-zero coefficients (with the coefficients for metal mining operations being significantly negative). Thus, there is no evidence that resource-rich regions experience faster growth of manufacturing; if anything, there is some evidence to the contrary.

In general, the results suggest that the most important factor that determines the spatial distribution of population (or industry) today is past population (or industry) distribution. Most likely, this is because attractive features of locations have manifested themselves through history: the best locations are the ones that are most densely settled. The spatial patterns of population and industry look very stable.

[14] In addition to these considerations, it is not always possible to correctly compare the coefficient values between time periods simply because different sets of variables are included in the equations for different time periods.

7.3 Projections for Russia

The difference between actual and counterfactual population levels (based on the main model) is mapped in Figure 5. The east-west divide is evident on the map. As a rule, the projected values of population in the western provinces are higher than the actual values, and they are lower than the actual in the eastern provinces. The degree of spatial autocorrelation is striking: *all* of the underdeveloped regions are located in the western part of the country. In the European part of the Former Soviet Union, there are only four observations with a predicted population that is distinctly lower than the actual, and one of them is Moscow.

In the West in general, provinces around larger cities (St. Petersburg, and also the capitals of the Union Republics, Kiev and Minsk) experience less of a population deficit in per capita terms. The population of Moscow is underpredicted. The rapid growth of Moscow in the 20th century comes from its being the capital of the Soviet Union. At the turn of the century, Moscow was only the second largest city in the Russian Empire.

The fact that Central Asian regions exhibit excess population can be explained by cultural factors: fertility rates in Central Asia and parts of the Caucasus are historically higher than in the rest of the country. I attempt to correct for these differences in Section 7.4.

The eastern regions – Siberia, and especially the Far East – are noticeably overpopulated, even though the predicted number for the Siberian population is very high. The main model (1) predicts about 19 million people. In fact in 1989 these provinces hosted more than 34 million people,[15], so there is an astonishing 14.5 million excess population east of the Urals. Moreover, the situation in the Urals is no better: they are also overpopulated.

Of course, the mere fact that the predicted population for individual regions appears to be over or under the actual level does not necessarily imply that the actual allocation is systematically distorted. First, the forecast model is not 100% precise. Second, the

[15] In 1989 the total population in the oblasts of Russian Soviet Federative Socialist Republic east of the Ural mountains was over 39 million. The difference is due to the fact that borders of imperial provinces I work with do not coincide with Soviet oblast borders. For example, the Russian city of Omsk, and most of modern Omsk oblast in the Russian Empire, belonged to Akmolinskaya province, a territory of which is now divided between Russian Federation and Kazakhstan. Because of that, I could not use Akmolinskaya province for my comparisons.

projected population levels for the individual regions are very sensitive to the location of the borders. Values for most of the regional characteristics are taken for the regional center, but population is spread throughout the territory.[16] If a large city is located near a regional border, then a small change in administrative borders may lead to large changes in the recorded regional population. Thus, even if Russian spatial allocation was produced by the same data-generating process that we used for Canada, individual regions might seem to be over- or underpopulated. However, we should expect that the differences between actual and predicted values for the neighboring regions approximately cancel each other out, and that the errors are spatially random, not systematic. Neither of these is true for Russia.

The main results are quite robust to the choice of model. The projected levels of population in the nine provinces of Siberia and the Far East for different models are reported in Table 3, row d).[17] Keep in mind that these results are likely biased towards Siberia, as the Canadian projected TPC levels suggest. Russian TPC is consistently predicted above actual, by 1.2 - 1.5°C (row i)).

Compared to the "naive" model (3), the Canada-based models predict more population in Russian Siberia. This is not surprising: Canadian spatial dynamics through the 20th century was characterized more by diffusion than concentration. The western parts of the country attracted population. The same growth through diffusion of population is predicted for the Russian periphery. Nonetheless, overpopulation of Siberia and the Far East emerges as a robust result, irrespective of model choice.

Results for manufacturing employment (row g)) are even more striking. Whereas population of the Russian east in the counterfactual world should be 30-40% lower, predicted manufacturing share on average is only 1/3 of actual.

To determine whether the difference between actual and predicted population of the nine provinces (row e)) is statistically different from zero, I conduct Monte-Carlo simulations. They estimate the probability that the existing population of 34 million in the eastern part of

[16] The fact that empirical results may depend on the specific configuration of the spatial unit borders is a known issue referred to as MAUP (modifiable aerial unit problem).

[17] On the map (Figure 3) they are numbered: 68(Tobolskaya), 69(Tomskaya), 63(Eniseiskaya), 65(Irkutskaya), 64(Zabaikalskaya), 62(Amurskaya), 70(Yakutskaya), 67(Primorskaya) and 68(Sakhalinskaya oblast).

the country indeed could be generated by the "Canadian" model. I draw 1,000 sets of random model coefficients (α's and β's), according to their estimated means and variance-covariance matrix. For each set of coefficients, I conduct the projection exercise on Russian/Soviet data and record the total projected population of the nine provinces. I then examine the sample of 1000 projections and record the distribution quantiles. This procedure is asymptotically equivalent to analytically evaluating the forecast error.

The lower 5% quantiles of the "excess" population in nine Russian provinces are reported in row f). In all of the models, the 5% quantile is far above zero. The difference between the actual and counterfactual populations is significant at a *very* high level of confidence.

7.4 Accounting for the WWII and fertility differences in the USSR

To what extent is the overpopulation of Siberia due to Soviet policy decisions versus exogenous factors, that is the circumstances beyond the control of Soviet authorities? The single most important historical event with an impact on the spatial pattern of Russia's economy is WWII. Any estimated overpopulation of Siberia and the Far East may be due to wartime destruction and evacuation of the western part of the country.

Furthermore, there is noticeable overpopulation in the regions of Central Asia which can be explained by traditionally high fertility rates in that area. Because it is important to disentangle the effects of culture from the effects of Soviet policy, I modify the projections procedure to try to account for these factors in this section.

WWII

WWII disproportionally affected the western regions of the country. The regions in the European part of the USSR suffered destruction of infrastructure and loss of many lives. In addition, a substantial number of strategically important enterprises, together with essential personnel, were evacuated to safer places during the war – mostly to the Urals, Siberia, and Central Asia.[18]

[18] "From July to November 1941, the equipment and machinery for more than 1,500 industrial enterprises (including 1,360 defense enterprises) were shipped eastward in 1.5 million train-car loads. To build and then stuff the Soviet defense plants, 10 million people – plant workers and their families – were relocated to the

If detailed information on population losses, infrastructure destruction, and evacuation efforts were available, it would be possible to account directly for the consequences of the war. Unfortunately, the lack of relevant data is a major obstacle. Detailed information on the evacuation efforts by the Soviet Government has not been published openly. Industry employment data at a low level of geographical aggregation were not published, even for peaceful times. The first post-war census of population in the Soviet Union took place in 1959. There is no way to obtain *oblast*-level data either on population loss or on loss of industrial capacity due to the war. Thus, instead of tracking the actual impact of the war, I try to construct an estimate of its long-run consequences.

According to the scattered evidence from various publications, Ukraine as a whole lost about 20% of its population and Belorussia lost about 25%.[19] The percentage loss of productive capabilities during WWII was not publicized in the Soviet Union, but from publications of gross industrial production relative to 1913 it can be inferred that actual production fell about 75% in the worst cases.[20] The Center and South of the European part of the Russian Federation were occupied for a shorter period of time, and people (as well as enterprises) had more time to evacuate. As the result, the loss of lives and productive capabilities was not as massive on average as in the westernmost regions of the Soviet Union.

To account for WWII, I take the projected population and industry values for the year 1941 and, instead of using them in the 1951 projections, I alter the values for those regions that were occupied during the war. To do so, I reduce the population levels of all affected regions by 25% and the industry employment levels by 75%. (That is, I am assuming that 75% of the productive infrastructure was destroyed and that 25% of the population was lost

East." – Gaddy (1996), p. 133.

[19]The following data were reported in statistical publications during the Soviet period: Belarus total population in 1939 was 8.9 million; losses of Belarus population during the war were more than 2.2 million people. Source: "Belorusskaya SSR za 20 let (1944-1963)" ("Belorussia during 20 years" – a statistical publication), Central Statistical Unit with the Government of Belorussian USR, "Belarus," Minsk 1964. Loss of only civilian population in Ukraine is reported as 16% of total. Source: "Ukraina za 50 rokiv" ("Ukraine during 50 years"), Central Statistical Unit with the Government of Ukrainian USR, Kiev 1967.

[20]In Leningrad region, the reported production levels for 1940 and 1945 correspondingly were 8.9 and 2.3 times higher than in 1913. The loss of production capabilities due to war, therefore, is about 74%. Similar figures are reported for Smolensk region. For Latvia and Estonia the losses are near 50%. Unfortunately, these data are not given for all regions. Source: "Atlas SSSR," Glavnoe upravlenie geodezii i kartografii, Moscow, 1962.

because of the war.)[21] These figures are deliberately higher than the actual losses. I then use these altered 1941 values for the 1951 projections. Equations (15) and (16) for 1951 become:

$$\ln Pop_{1951}^{RusF} = \beta_{1951}^{Rus} + \hat{\beta}_{1951}^{pop} \ln(0.75^{w} Pop_{1941}^{RusF}) + \hat{\beta}_{1951}^{ind} \ln(0.25^{w} Ind_{1941}^{RusF}) + \sum_{k} \hat{\beta}_{1951}^{k} x_{1941}^{k}, \tag{17}$$

$$\ln Ind_{1951}^{RusF} = \alpha_{1951}^{Rus} + \hat{\alpha}_{1951}^{pop} \ln(0.75^{w} Pop_{1941}^{RusF}) + \hat{\alpha}_{1951}^{ind} \ln(0.25^{w} Ind_{1941}^{RusF}) + \sum_{k} \hat{\alpha}_{1951}^{k} x_{1941}^{k}, \tag{18}$$

where w is a dummy indicator if the region was occupied during WWII. The rest of the process is unchanged.

Implicitly embedded in this procedure is the assumption that any shock due to the war has to be permanent. Because the coefficients of the dynamic relationship for the years after 1941 were not changed, I am imposing an equilibrium path onto the Russian economy which had been shaken by a major shock. The results of this procedure overdramatize the situation and overestimate the effect of war on a counterfactual market economy. From the work of Davis & Weinstein (2002) and Brakman, Garretsen & Schramm (2004), we know that war shocks tend to be transitory: people tend to rebuild destroyed cities and population levels tend to rebound.[22] Even in the absence of pure economic incentives, people tend to return home, even if home was destroyed. In my WWII counterfactual this effect is completely ignored.

On the other hand, given Russia's size and higher transportation costs, it is possible that the WWII shock might have had more severe long-term consequences. In Japan, people moved out of cities to the countryside when running from war, but in Russia people moved across the country. Thus it would be more costly for the Russians to move back to the previous spatial allocation once the war was over than it was for the Japanese. Probably,

[21] It has to be noted that due to the relative nature of the multinomial logit model, reducing the (population or industry) shares of one region automatically raises shares of the regions unaffected. Thus, the composite effect of this artificial shock is even larger than nominal 25% and 75%.

[22] Interestingly, mean-reversion after the war shocks is not observed in socialist countries: Eastern Germany and USSR. In the Soviet case, it is quite consistent with the policy: the most of the evacuated defense enterprises not be moved back.

neither the projection with zero-shock nor the projection with fully permanent shock will produce the perfect counterfactual allocation: the truth lies somewhere in between. The projections corrected for WWII therefore should be treated as an upper bound of the long-run consequences of the war.[23]

Figure 6 presents the results of the projections.[24] With the artificial shock included, the difference between actual and projected population in Siberia and the Far East decreases substantially: it falls to 9.6 million. Thus, the long-term effect of the war generously allows for about 5 million more people in Siberia. However, the difference between actual and projected population still remains statistically significant (at about the 99.5% level, according to Monte-Carlo simulations)! Even with the (probably grossly exaggerated) war effect built in, the estimated excess of population in Siberia and the Far East is too large to be generated only by random error; it has to be the result of the deliberate policy by the Soviet authorities.

Fertility

In the late USSR fertility rates varied from 1.93 children per woman average in Estonia to 5.03 in Turkmenistan. If labor mobility were perfect, then the birth rate differences would not affect the spatial distribution of the population in long-run: people would instantly re-allocate (migrate) according to the economic incentives alone. This is not the case if mobility is imperfect. For example, if people are more likely to migrate to parts of the country with similar culture and/or ethnic composition, then the differences in fertility rates can have a long-term effect on spatial population structure.

Obviously, ignoring such differences would skew the population projection results away from the Central Asian republics and toward regions with lower natural population growth. Historically, Russia had one of the lowest birth rates in the former USSR. Siberia had roughly

[23]Putting it a different way, in a reduced-form model the past levels of population carry the effect of inherent location quality. Applying the shock to the reduced-form model, I am implicitly treating the regions that suffered from war and lost population as if they were inherently undesirable (and therefore less populated). In this case, the forecast from 1941 on is biased toward regions that were spared in war, i.e. it exaggerates the war's impact.

[24]The results are not particularly sensitive to magnitude of the shock. I used the range of percentage decrease values, from 20% to 50% for population, and from 60% to 100% for industry. Only at the level of 50% loss of population and 100% loss of industry did I get a projected population for the Eastern regions close to the actual level.

the same (age-standardized) low birth rates as the European part of the Russian Federation. Thus, correcting for fertility differences on average would decrease the estimated population deficit in the European part of the country and increase the estimated population surplus in Siberia and the Far East.

To account for fertility differences, and before plugging the past population levels into the projection equations, I multiply them by the population growth coefficients that are specific to a geographical location and time period. Equations (15) and (16) now become:

$$\ln Pop_{1921}^{RusF} = \beta_{1921}^{Rus} + \hat{\beta}_{1921}^{pop} \ln((1+g)^{10} Pop_{1911}) + \hat{\beta}_{1921}^{ind} \ln Ind_{1911} + \sum_k \hat{\beta}_{1921}^{k} x_{1911}^{k}, \quad (19)$$

$$\ln Ind_{1921}^{RusF} = \alpha_{1921}^{Rus} + \hat{\alpha}_{1921}^{pop} \ln((1+g)^{10} Pop_{1911}) + \hat{\alpha}_{1921}^{ind} \ln Ind_{1911} + \sum_k \hat{\alpha}_{1921}^{k} x_{1911}^{k}, \quad (20)$$

where g is the annual natural population growth rate. The data on birth rates in the USSR are consistently available only at the level of Union Republics, and for 1940 and then from 1960 on. Thus, I had to use the 1940 growth rates for the 1921-to-1951 projections.

The same critique applies here as with the WWII procedure: I am introducing exogenous changes into the reduced-form relationship. The regions with higher population levels due to high birth rates, and the regions that are historically more attractive to migration, are going to be treated the same way. This procedure over-predicts the growth of high-fertility regions. Thus, the results of this exercise are an upper bound on what could be the result of fertility (and mortality) differentials.

The projection results are shown in Appendix B in Tables 7 and 8, and on Figures 7 (without WWII simulations) and 8 (with WWII simulations). The summary of the modified projections for the main model and for the alternatives is in Table 2.

7.5 Temperature per capita dynamics

The trajectories of the actual and projected TPC in Canada and in Russia are shown in Figures 9 and 10. The plot of the Canadian TPC trajectory (Figure 9) shows that forecast models are somewhat "colder" than reality. This is probably an artifact of the large negative

Model	Excess population in 9 provinces, thousand			
	No correction	Correction for WWII	Correction for fertility	Correction for WWII and fertility
Main result				
min. relative SSE (1)	14 576 (9 783)	9 563 (3 676)	17 583 (13 036)	14 012 (8 760)
Alternative models				
min. absolute SSE (2)	17 291 (13 359)	12 471 (6 199)	19 677 (14 876)	16 172 (10 553)
Akaike (8)	13 731 (8 812)	5 680[a] (-531)	16 415 (12 246)	10 403 (5 098)

lower 5% simulated quantile in parentheses;
[a] - significant at 10% level, but not on 5% level

Table 2: Excess population in 9 provinces of Siberia and the Far East, adjusted projections.

error for several major cities. The population of Vancouver is severely underestimated; the Vancouver error alone is enough to account for the 0.15 degree actual-projected TPC difference. If this bias were corrected, the forecast for Russia would be even “warmer”.

Generally, the trajectory of counterfactual Russian TPC mirrors the Canadian dynamics. In Canada, TPC dips around the years 1920-1940, then rises steadily until 1990. At the beginning of the century, after the construction of the Trans-Canadian Railroad, settlers rushed to Alberta and Saskatchewan, drawn by an abundance of fertile land. Aggregate temperature dropped slightly as the colder areas were populated. Later, as agricultural technologies became less labor intensive, and as the share of agriculture in Canadian GDP fell, the population shifted towards manufacturing centers, away from agricultural areas, and the TPC went up.

At least by TPC dynamics, similar processes were taking place in the late Russian Empire - early USSR. Peasant migration to Siberia of the beginning of 20th century, depopulation of major cities during the October revolution, and civil war together led to comparable growth of the frontier regions in Russia. Up until 1940, actual and counterfactual trajectories for Russia were practically the same. A noticeable divergence begins after WWII, when the counterfactual model predicts a TPC “reversal” *a la* Canada. The actual Russia never started to "warm up." While in Canada (and in the counterfactual Russian model) after the 1930s the productive resources became concentrated in regions already established as manufacturing and service centers, for Soviet Russia there was instead continued development of its frontier.

In Figure 10 the shaded area represents the 95% confidence interval around the predicted TPC trajectory (without corrections for either fertility or war). The permanence of the artificial WWII shock is also evident from the graph: the TPC trajectories "with" and "without" war do not converge; the gap widens over time. Although grossly exaggerated, the effect of WWII still does not explain the entire actual-predicted TPC gap. Actual TPC remains well below all of the possible counterfactual estimates.

8 Conclusions

I show that the present allocation of population and industry in Russia inherited from the Soviet system is far different from what would have occurred in the absence of Soviet location policy: it is colder and further to the east. The Eastern part of the country is overpopulated by about 14.5 million people compared to the counterfactual market allocation. This result is robust to the model selection.

The impact of WWII, however drastic in the case of Russia, only partly explains the east-west imbalance. Even according to the most liberal estimates and with the war adjustment, excess population in Siberia and the Far East remains at a level above 9.6 million and is statistically significant.

With the transition to a market economy, and as agents have been able to freely respond to market stimuli, it might have been hoped that the spatial inefficiency would eventually correct itself, that is, that people would migrate to more favorable places. Indeed, beginning in the early 1990s internal migration patterns reversed in post-communist Russia. People started to leave the North, the Far East, and Eastern Siberia. Two of the most remote territories with unbearable climate – Magadan *oblast* and Chukotka republic – lost 30-40% of population during the transition years. Still, even at that uncommonly high rate of out-migration, it would have taken 50-60 years at a minimum to revert to the counterfactual population levels in these territories.

By the end of 1990s, this return migration outflow from the "Far North" gradually decreased to a minimal level. It is now clear that the reversal of Soviet location policies will not happen in the foreseeable future. At the current migration rate, a return to the counterfactual spatial allocation would take about 180 years. Thus, the costs of spatial inefficiency will be an extra burden for Russian Federation for years to come.

References

Acemoglu, D., Johnson, S. & Robinson, J. A. (2001), 'The Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review* **91**(5), 1369–1401.

Andrienko, Y. & Guriev, S. M. (2004), 'Determinants of Interregional Mobility in Russia', *Economics of Transition* **12**(1), 1–27.

Arauzo-Carod, J.-M., Liviano-Solis, D. & Manjón-Antolín, M. (2010), 'Empirical Studies In Industrial Location: An Assessment Of Their Methods And Results', *Journal of Regional Science* **50**(3), 685–711.

Bloom, D. & Sachs, J. D. (1998), 'Geography, Demography and Growth in Africa', *Brookings Papers on Economic Activity* **2**.

Brakman, S., Garretsen, H. & Schramm, M. (2004), 'The Strategic Bombing of German Cities during World War II and its Impact on City Growth', *Journal of Economic Geography* **4**(2), 201–218.

Carlton, D. W. (1983), 'The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables', *The Review of Economics and Statistics* **65**, 440–449.

Crozet, M. (2004), 'Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model', *Journal of Economic Geography* **4**(4), 439–458.

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

Dienes, L. (1972), 'Investment Priorities in Soviet Regions', *Annals of the Association of American Geographers* **62**(3), 437–454.

Ellison, G. & Glaeser, E. L. (1997), 'Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach', *Journal of Political Economy* **105**, 889–927.

Gaddy, C. G. (1996), *The Price of the Past: Russia's Struggle with the Legacy of a Militarized Economy*, Brookings Institution Press, Washington, D.C.

Gaddy, C. G. & Ickes, B. W. (2001), 'The Cost of the Cold', The Pennsylvania State University working paper, unpublished.

Gallup, J. L., Gaviria, A. & Lora, E. (2003), *Is Geography Destiny?*, A copublication of Stanford Social Sciences, an imprint of Stanford University Press, and the World Bank, Palo Alto, CA and Washington, DC.

Gallup, J. L., Sachs, J. D. & Mellinger, A. (1999), 'Geography and Economic Development', *International Regional Science Review* **22(2)**, 179–232.

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

Head, K. & Mayer, T. (2004), 'Market Potential and the Location of Japanese Investment in the European Union', *The Review of Economics and Statistics* **86**(4), 959–972.

Hill, F. & Gaddy, C. (2003), *The Siberian Curse. How Communist Planners Left Russia Out in the Cold.*, Brookings Institution Press, Washington, DC.

Islam, N. (1995), 'Growth Empirics: A Panel Data Approach', *Quarterly Journal of Economics* **110**.

Lynch, A. C. (2002), 'Roots of Russia's Economic Dilemmas: Liberal Economics and Illiberal Geography', *Europe-Asia Studies* **154**, 31–49.

McFadden, D. (1974), Conditional Logit Analysis of Qualitative Choice Behavior, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic press: New York, pp. 105–142.

Ozornoy, G. I. (1991), 'Some Issues of Regional Inequality in the USSR under Gorbachev', *Regional Studies* **25**(5), 381 393.

Parshev, A. P. (1999), *Why Russia is not America*, Forum, Moscow.

Rappaport, J. & Sachs, J. D. (2001), 'US as a Coastal Nation', Research Division, Federal Researve Bank of Kansas City.

Redding, S. & Venables, A. J. (2004), 'Economic geography and international inequality', *Journal of International Economics* **62**(1), 53–82.

Rodgers, A. (1974), 'The Locational Dynamics Of Soviet Industry', *Annals of the Association of American Geographers* **64**(2), 226–240.

A The search algorithm for choosing the best model

The algorithm for choosing the best model works as follows:

Step 1 Choose the "core" set of variables – i.e. the ones that definitely are going to be included into the model. In our case, the past values of population and industry, area and the constant term have to be included. This narrows down the number of variables "in question" from 224 to 168.

Step 2 Start with the model with all variables included. Drop one of the 168 "questionable" variables from the regressions and estimate the restricted model. Perform a forecast on Canadian data and evaluate it according to a chosen criterion. If the forecast is better without the variable - drop it, if worse - keep it. Repeat the procedure for all remaining non-core variables consecutively.

Step 3 Take the model that resulted from step 2. Now try to add explanatory variables and check if inclusion of any of them improves the forecast. If it does, put a variable back into regressions.

Step 4 Repeat steps 2 and 3 until no more inclusions or exclusions can be made that would improve the dynamic forecast. A local minimum for the chosen criterion is found.

Procedure necessarily converges, since at every step the value of SSE is decreased. Of course, there is no guarantee that the procedure finds the global minimum of SSE. If there exist several local minima that correspond to the different non-nested models, it is possible that algorithm finds one of those models, not necessarily the best one.

Step 5 Change the order of "questionable" variables and repeat steps 2 to 4, examining the variables "in question" in different order. If the procedure finds a different local maximum (a different set of variables), compare it with the one found previously and pick the better one. Repeat several times.

Repeat steps 2 to 4, but with different starting point. For example, start with the model that includes "core" variables only. As there are no more variables to drop, go directly to step 3, then do as algorithm requires. If a different model results, compare it with the one found previously and pick the best. Repeat with various starting points. If different local minima are found, select the one with lowest SSE. [25]

B Tables and figures

[25] Step 5 did not uncover a better alternative, i.e. in this case the minimum found in Step 4 is likely global.

Model	**(1) min. rel. SSE**	(2) min. abs. SSE	(3) 1911 extrapolated	(4) fixed over time	(5) parsimonious	(6) geography only	(7) full set	(8) Akaike	Actual
Performance on Canadian data									
a) R^2 for absolute errors (population, 1991)	**0.921**	0.941	0.8660	0.872	0.861	0.894	0.830	0.827	
b) R^2 for relative errors (population, 1991)	**0.883**	0.846	0.514	0.875	0.589	0.819	0.860	0.787	
c) TPC, Canada 1991	**-9.038**	-8.749	-9.476	-9.335	-9.462	-9.194	-9.291	-9.333	-8.563
d) Moran's I for forecast residuals (p-value)	**0.093 (0.139)**	0.114 (0.096)	0.536* (0.000)	0.006 (0.380)	0.555* (0.000)	0.138 (0.064)	-0.010 (0.438)	0.083 (0.155)	
Projections for Russia									
e) Population in 9 provinces of Siberia and the Far East, 1991	**19 672**	16 957	14 862	17 567	15 205	23 243	21 218	20 517	34 248
f) Excess population in 9 provinces, 1991	**14 576****	17 291*	19 386	16 681*	19 043*	11 005*	13 030*	13 731*	
g) Lower 5% simulated quantile of the excess pop. in 9 provinces	**9 783**	13 359	-	12 106	17 891	7 094	6 253	8 812	
h) Share of manuf. employment in 9 provinces out of RF total, 1991	**5.8%**	3.0%	7.9%	3.0%	8.3%	11.5%	6.7%	8.3%	20%
i) TPC Russian Federation, 1991	**-11.428**	-11.029	-11.446	-11.150	-11.457	-11.801	-11.552	-11.632	-12.944

*different from zero at 95% confidence level.

Table 3: Projection models

Dependent Variable (ln)	*POP* 1921	*POP* 1931	*POP* 1941	*POP* 1951	*POP* 1961	*POP* 1971	*POP* 1981	*POP* 1991
Population_{t-1}	1.097*	0.997*	0.982*	0.811*	0.861*	0.861*	0.957*	1.076*
	(0.045)	(0.035)	(0.035)	(0.041)	(0.044)	(0.052)	(0.058)	(0.045)
$\text{Manufacturing}_{t-1}$	-0.143*	-0.007	0.001	0.117*	0.093*	0.079*	0.010	-0.041
	(0.033)	(0.030)	(0.022)	(0.036)	(0.034)	(0.034)	(0.043)	(0.031)
Area	0.048*	0.055*	0.031*	0.045*	0.048*	0.066*	0.043*	0.011
	(0.016)	(0.015)	(0.017)	(0.016)	(0.012)	(0.015)	(0.019)	(0.013)
Temperature	0.007			0.009*	0.001	0.008*	0.012*	0.000
	(0.007)			(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Distance to Toronto		0.000	-0.066*				0.008	-0.011
		(0.000)	(0.026)				(0.025)	(0.023)
Railroads_{t-1}	0.030*	-0.010	-0.005	-0.010				-0.014
	(0.017)	(0.013)	(0.012)	(0.011)				(0.013)
Coal mining_{t-1}	0.043		0.002			0.029		
	(0.046)		(0.043)			(0.043)		
$\text{Metals mining}_{t-1}$	-0.021	0.005	0.113*	-0.020			-0.046	
	(0.055)	(0.054)	(0.055)	(0.047)			(0.049)	
$\text{Oil extraction}_{t-1}$		-0.042	-0.060		0.092*	-0.036	0.087*	
		(0.076)	(0.068)		(0.056)	(0.054)	(0.044)	
Timber cutting	0.150*	0.030			-0.005			
	(0.049)	(0.048)			(0.038)			
Port	-0.043	-0.050		-0.040	-0.035	-0.069	-0.054	-0.016
	(0.041)	(0.041)		(0.038)	(0.039)	(0.041)	(0.040)	(0.039)
Trade route	-0.016	0.010	-0.014			-0.065	-0.072	-0.029
	(0.045)	(0.049)	(0.041)			(0.042)	(0.043)	(0.043)
Agricultural land			-0.035			0.060		
			(0.051)			(0.044)		
Urbanization, 1911	0.212	0.343*		0.172	0.263*	0.276*	0.251*	0.228*
	(0.142)	(0.172)		(0.121)	(0.096)	(0.129)	(0.136)	(0.139)

Number of observations=279, parameters = 77, R^2=0.99.

Standard errors are in parentheses. * indicates significance at 90% level.

Table 4: Estimation results for the main model. Equations for population.

Dependent Variable (ln)	IND 1941	IND 1951	IND 1961	IND 1971	IND 1981	IND 1991
Population_{t-1}	0.735*	0.070	-0.077	0.139	-0.089	0.552*
	(0.077)	(0.080)	(0.096)	(0.114)	(0.124)	(0.095)
$\text{Manufacturing}_{t-1}$	0.248*	0.938*	1.095*	0.875*	1.173*	0.509*
	(0.073)	(0.068)	(0.073)	(0.075)	(0.085)	(0.059)
Area	0.167*	0.041	0.095*	0.041	0.105*	-0.062*
	(0.040)	(0.036)	(0.037)	(0.034)	(0.037)	(0.024)
Temperature	0.030*	0.033*	0.019*			
	(0.011)	(0.009)	(0.010)			
Distance to Toronto	-0.326*			-0.006	0.231*	-0.206*
	(0.067)			(0.061)	(0.051)	(0.047)
Railroads_{t-1}	0.075*	0.020	0.065*	-0.013		
	(0.030)	(0.026)	(0.024)	(0.026)		
Coal mining_{t-1}	0.045		0.040	0.106	-0.067	
	(0.101)		(0.091)	(0.105)	(0.092)	
$\text{Metals mining}_{t-1}$		-0.315*	-0.211*		-0.402*	
		(0.128)	(0.107)		(0.101)	
$\text{Oil extraction}_{t-1}$	-0.004		0.010	-0.207	0.014	
	(0.158)		(0.131)	(0.133)	(0.111)	
Timber cutting	-0.022	0.200*		-0.183*	-0.059	
	(0.109)	(0.085)		(0.097)	(0.098)	
Port	-0.093	-0.128		-0.038*		
	(0.094)	(0.083)		(0.089)		
Trade route	-0.061		-0.131	-0.076		-0.200*
	(0.103)		(0.073)	(0.101)		(0.090)
Agricultural land	-0.113	-0.199		0.051	-0.275*	
	(0.099)	(0.111)		(0.100)	(0.091)	
Urbanization, 1911	1.435*			0.091		0.561*
	(0.353)			(0.316)		(0.269)

Number of observations=180, parameters=58, R^2=0.98

Standard errors are in parentheses. * indicates significance at 90% level.

Table 5: Estimation results for the main model. Equations for manufacturing.

Region	Projected population, thousands	Actual population, thousands	Difference, thousands	Actual to projected ratio
Alberta north	1115.9	1268.6	152.7	1.14
Alberta central	1021.2	1063.5	42.3	1.04
Alberta south	234.0	213.4	-20.6	0.91
BC coast	965.8	679.6	-286.2	0.70
Kootenay	314.2	176.3	-137.9	0.56
Vancouver area	1174.7	1833.0	658.3	1.56
Yale and Cariboo	455.0	593.2	138.2	1.30
Manitoba center	126.1	80.5	-45.6	0.64
Manitoba sw	109.2	47.9	-61.2	0.44
Manitoba south-center	253.7	152.0	-101.7	0.60
Manitoba south-east	786.9	747.3	-39.5	0.95
Manitoba north	61.1	64.2	3.1	1.05
New Brunswick north west	179.8	165.7	-14.1	0.92
New Brunswick south	411.3	346.7	-64.6	0.84
New Brunswick east coast	186.3	211.5	25.2	1.14
Nova Scotia west	901.5	657.7	-243.9	0.73
Nova Scotia east	594.2	242.3	-351.9	0.41
Toronto area	4564.4	5897.1	1332.7	1.29
Ontario south	1594.8	1235.9	-358.9	0.77
Ottawa area	1493.0	1219.0	-274.0	0.82
Ontario center	1177.8	910.5	-267.3	0.77
Ontario north	925.2	581.9	-343.3	0.63
Ontario north-west	601.3	240.6	-360.7	0.40
PEI	120.7	129.8	9.1	1.08
Montreal and around	2245.6	3442.3	1196.7	1.53
Quebec south	752.5	861.5	108.9	1.14
Quebec city and around	596.0	842.8	246.8	1.41
Quebec center	415.4	377.8	-37.5	0.91
Quebec east	430.8	331.0	-99.8	0.77
Quebec north east	616.7	417.0	-199.7	0.68
Quebec west	141.5	203.4	61.9	1.44
Quebec south-west	211.0	332.3	121.3	1.57
Saskatchewan south	786.4	426.4	-360.1	0.54
Saskatchewan center	366.8	337.8	-29.0	0.92
Saskatchewan north	352.6	224.8	-127.9	0.64
Yukon	26.9	27.8	0.9	1.03
N-W Territ	37.3	57.6	20.4	1.55
Newfoundland	861.5	568.5	-293.0	0.66

Table 6: Projected vs actual population. Canada, 1991. Main model (min relative SSE).

	Province	Actual population	Projected population	Projected population, WWII adjusted	Projected population, fertility adjusted	Projected population, WWII and fertility adjusted
			European provinces			
1	Arkhangelskaya	2699.1	1088.1	1365.4	921.8	1119.3
2	Astrakhanskaya	1426.5	2383.9	2991.5	2019.6	2452.4
3	Bessarabskaya	4345.8	3682.0	2866.9	3615.5	2724.1
4	Vilenskaya	1977.2	2868.3	2233.3	2363.7	1780.9
5	Vitebskaya	2868.4	3628.2	2825.0	3443.3	2594.4
6	Vladimirskaya	1089.7	3350.7	4204.6	2838.5	3446.8
7	Vologodskaya	1970.2	5544.7	6957.9	4697.3	5703.9
8	Volynskaya	4037.9	5362.0	4175.1	4490.3	3383.3
9	Voronezhskaya	2734.1	3989.2	3106.1	3379.4	2546.3
10	Vyatskaya	3431.5	5469.3	6863.3	4633.4	5626.4
11	Grodnenskaya	1394.9	2991.8	2329.5	2839.3	2139.3
12	Oblast' Voiska Donskogo	5424.8	5918.4	4608.3	5013.9	3777.8
13	Ekaterinoslavskaya	11286.0	4856.2	3781.2	4066.7	3064.1
14	Kazanskaya	3806.9	3860.2	4844.0	3270.2	3971.0
15	Kaluzhskaya	1066.8	2248.7	1750.9	1905.0	1435.4
16	Kievskaya	6170.8	7808.5	6080.0	6539.1	4927.0
17	Kovenskaya	2314.9	3145.7	2449.4	2592.3	1953.2
18	Kostromskaya	1425.2	2901.5	3641.0	2458.0	2984.8
19	Kurlyandskaya	917.0	1078.9	840.1	613.1	462.0
20	Kurskaya	2564.5	3452.7	2688.4	2925.0	2203.9
21	Liftlyandskaya	2006.3	2406.3	1873.7	1367.5	1030.4
22	Minskaya	4601.6	5890.1	4586.2	5589.9	4211.8
23	Mogilevskaya	2145.7	3409.6	2654.8	3235.8	2438.1
24	Moskovskaya	15682.4	12044.3	9378.1	10203.4	7687.9
25	Nizhegorodskaya	3336.9	2863.6	3593.5	2426.0	2945.8
26	Novgorodskaya	1454.6	3230.7	2515.5	2736.9	2062.2
27	Olonetskaya	892.4	907.2	706.4	768.6	579.1
28	Orenburgckaya	6421.7	3683.1	4621.8	3120.2	3788.9
29	Orlovskaya	2084.7	3534.4	2752.0	2994.2	2256.0
30	Penzenskaya	1773.2	1983.9	2489.5	1680.7	2040.8
31	Permskaya	7755.0	7153.5	8976.7	6060.1	7358.8
32	Podol'skaya	2660.9	5155.3	4014.1	4317.2	3252.9
33	Poltavskaya	2487.5	4866.4	3789.1	4075.3	3070.6
34	Pskovskaya	771.7	2377.6	1851.3	2014.2	1517.6
35	Ryazanskaya	1432.0	3592.5	4508.2	3043.4	3695.7
36	Samarskaya	4267.3	5170.0	6487.6	4379.8	5318.4
37	St. Peterburgskaya	6446.6	7855.9	6116.9	6655.2	5014.5
38	Saratovskaya	2117.7	4562.3	5725.1	3865.0	4693.3
39	Simbirskaya	2441.0	2191.4	2749.9	1856.4	2254.3
40	Smolenskaya	1211.2	3668.7	2856.6	3108.0	2341.7
41	Tavricheskaya	4916.7	3263.3	2540.9	2732.8	2059.1
42	Tambovskaya	2502.7	4294.4	5389.0	3638.1	4417.7
43	Tverskaya	1581.4	4129.1	3215.1	3498.0	2635.6
44	Tul'skaya	1891.4	2825.5	2200.0	2393.6	1803.5
45	Ufimskaya	4533.7	3990.1	5007.1	3380.3	4104.7

Table 7: Projected vs actual population in imperial province borders, according to the main model (min relative SSE). The Soviet Union, 1989 (part 1).

	Province	Actual population	Projected population	Projected population, WWII adjusted	Projected population, fertility adjusted	Projected population, WWII and fertility adjusted
			European provinces, cont.			
46	Khar'kovskaya	4830.8	5352.8	4167.8	4482.6	3377.4
47	Khersonskaya	4524.1	4397.6	3424.1	3682.7	2774.8
48	Chernigovskaya	2317.2	4715.1	3671.3	3948.6	2975.1
49	Estlyandskaya	1037.9	487.7	379.7	264.2	199.1
50	Yaroslavskaya	1384.1	2207.7	2770.4	1870.3	2271.1
			Caucasus			
51	Bakinskaya	4158.6	1674.7	2101.5	2761.1	3352.8
52	Batumskaya	392.7	133.3	167.2	156.6	190.2
53	Dagestanskaya	1514.1	864.3	1084.5	1424.9	1730.3
54	Elisavetpol'skaya	2536.2	2301.0	2887.4	3793.7	4606.7
55	Kubanskaya	4929.3	6013.9	4682.7	5094.8	3838.7
56	Kutaisskaya	1441.9	1536.8	1928.5	1806.1	2193.1
57	Stavropol'skaya	1685.1	1986.0	1546.3	1682.4	1267.6
58	Terskaya	4209.6	3367.4	2622.0	5551.9	4183.2
59	Tiflisskaya	2898.9	3456.8	4337.8	4062.4	4933.0
60	Chernomorskaya	585.7	157.3	197.4	133.3	161.9
61	Erivanskaya	3582.8	1715.6	2152.9	3498.5	4248.3
			Siberia and the Far East			
62	Amurskaya	1333.4	498.0	625.0	421.9	512.3
63	Eniseiskaya	4244.4	3234.7	4059.2	2740.3	3327.6
64	Zabaikal'skaya	2496.1	1180.7	1481.7	1000.3	1214.6
65	Irkutskaya	2966.9	1675.9	2103.0	1419.7	1724.0
66	Primorskaya	5009.5	731.9	918.4	620.0	752.9
67	Sakhalinskaya	709.6	22.5	28.2	19.0	23.1
68	Tobol'skaya	6245.8	5688.5	7138.4	4819.1	5851.8
69	Tomskaya	10160.8	6112.0	7669.8	5177.8	6287.5
70	Yakutskaya	1081.4	527.3	661.6	446.7	542.4
			Central Asia			
71	Akmolinskaya	5449.1	2272.0	2851.0	3611.3	4385.3
72	Zakaspiiskaya	2424.1	1765.1	2214.9	3810.5	4627.0
73	Samarkandskaya	2290.9	2181.8	2737.9	5060.4	6144.8
74	Semipalatinskaya	2576.7	1866.8	2342.6	2967.4	3603.3
75	Semirechenskaya	6489.2	2211.0	2774.5	3920.6	4760.8
76	Syr-Dar'inskaya	7971.0	6433.8	8073.5	14922.1	18120.0
77	Turgaiskaya	1629.8	1522.6	1910.7	2420.2	2938.9
78	Ural'skaya	1102.0	2125.1	2666.7	3377.8	4101.7
79	Ferganskaya	10466.4	5948.8	7465.0	12310.0	14948.0

Table 8: Projected vs actual population in imperial province borders, according to the main model (min relative SSE). The Soviet Union, 1989 (part 2).

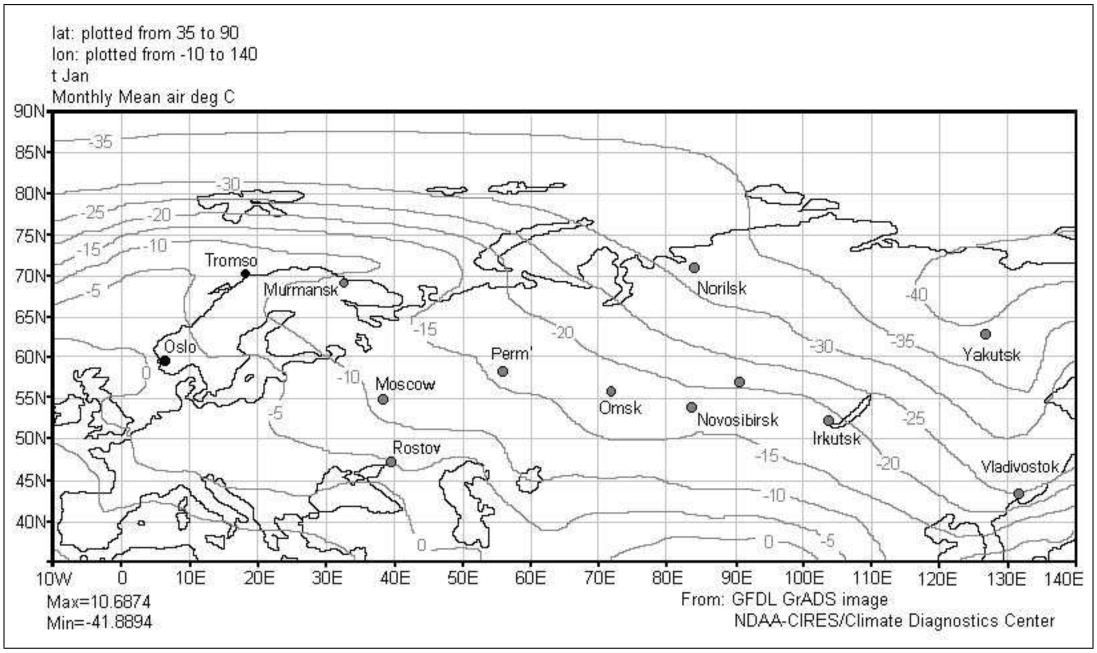


Figure 1: Isotherms: average January air temperature.

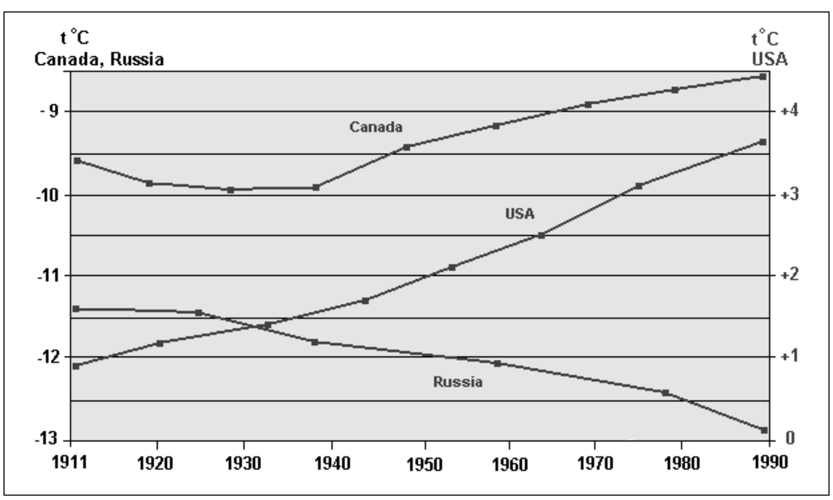


Figure 2: Change in TPC index in Russia, Canada (left scale) and USA (right scale).

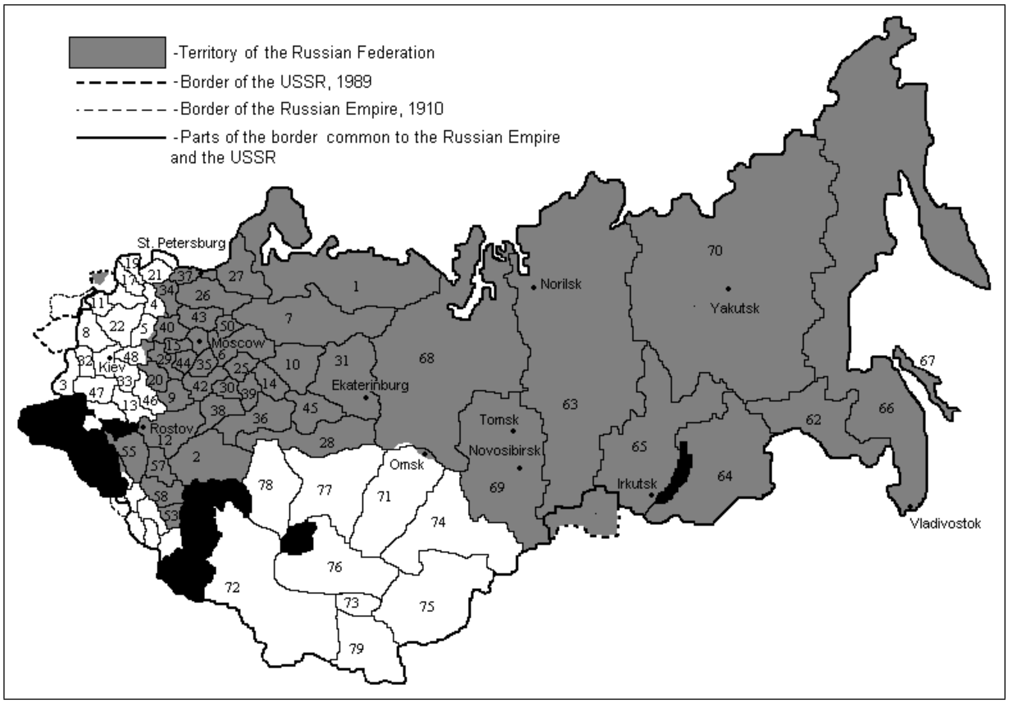


Figure 3: Administrative divisions in the Russian Empire and the borders of the Soviet Union and Russian Federation.

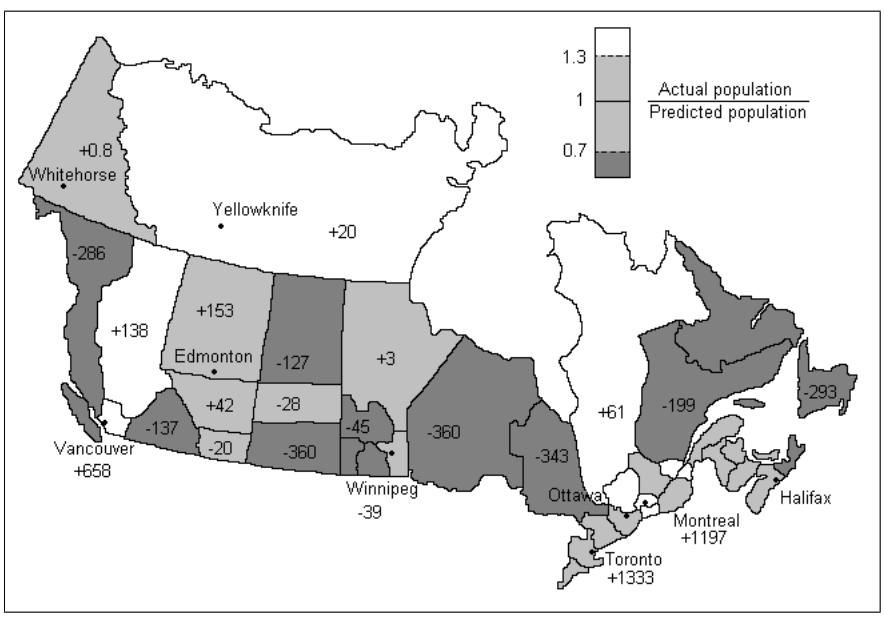


Figure 4: Projected vs. actual population according to the main model, Canada 1991. (Numbers show the absolute difference between projected and actual population, in thousands).

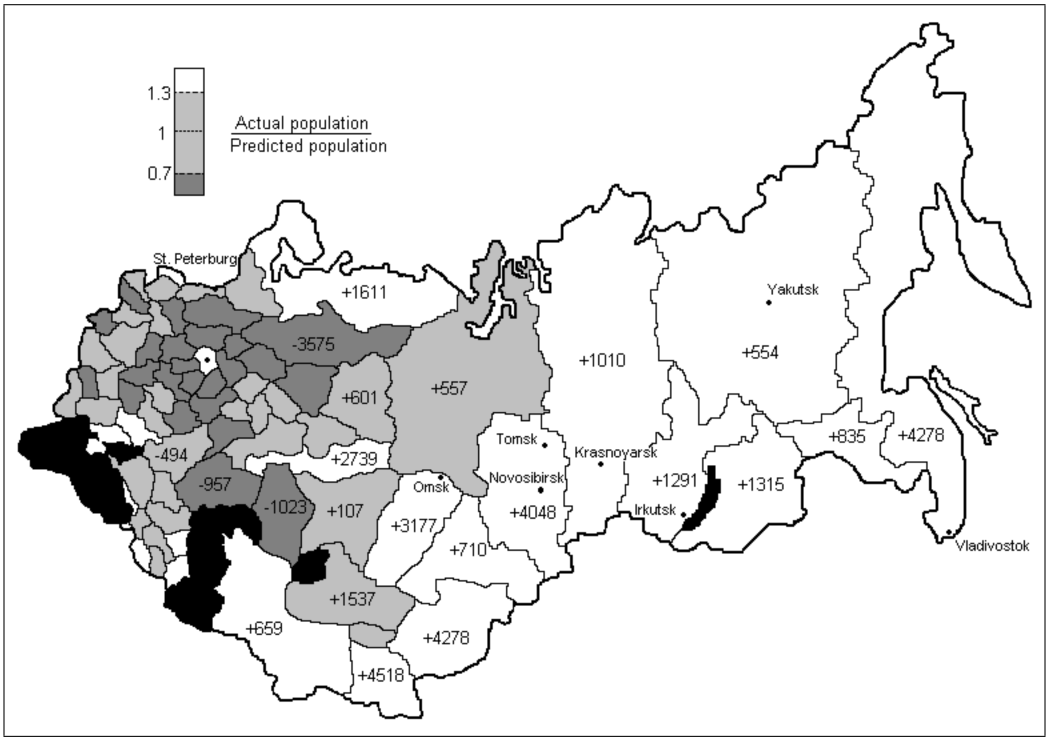
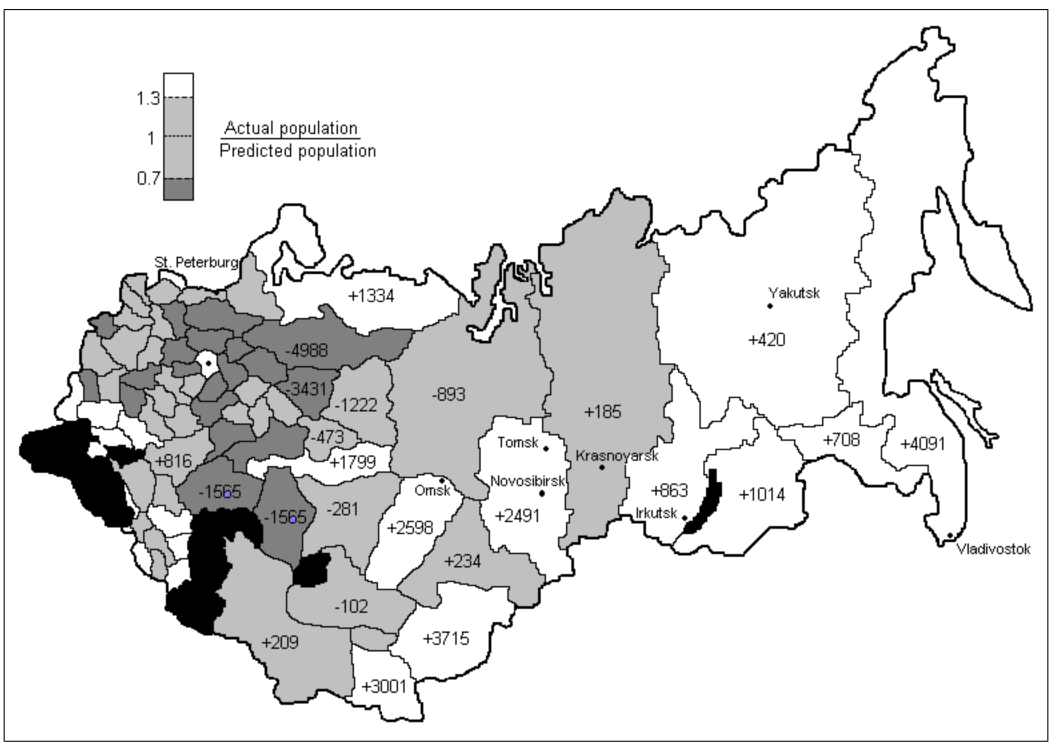


Figure 5: Projected vs. actual population according to the main model, USSR 1991. (Numbers show the absolute difference between projected and actual population, in thousands).



Numbers show the absolute difference between projected and actual population, in thousands.

Figure 6: Projected vs. actual population, USSR. Accounted for WWII.

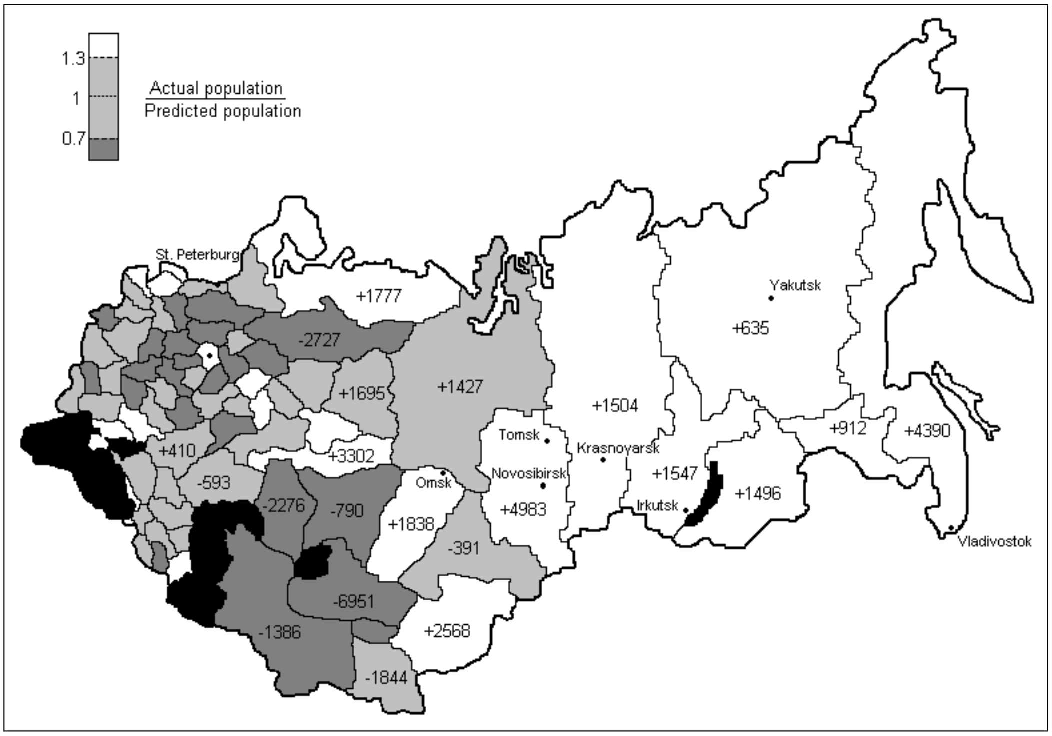


Figure 7: Projected vs. actual population, USSR. Corrected for fertility differences (numbers show the absolute difference between projected and actual population, in thousands.

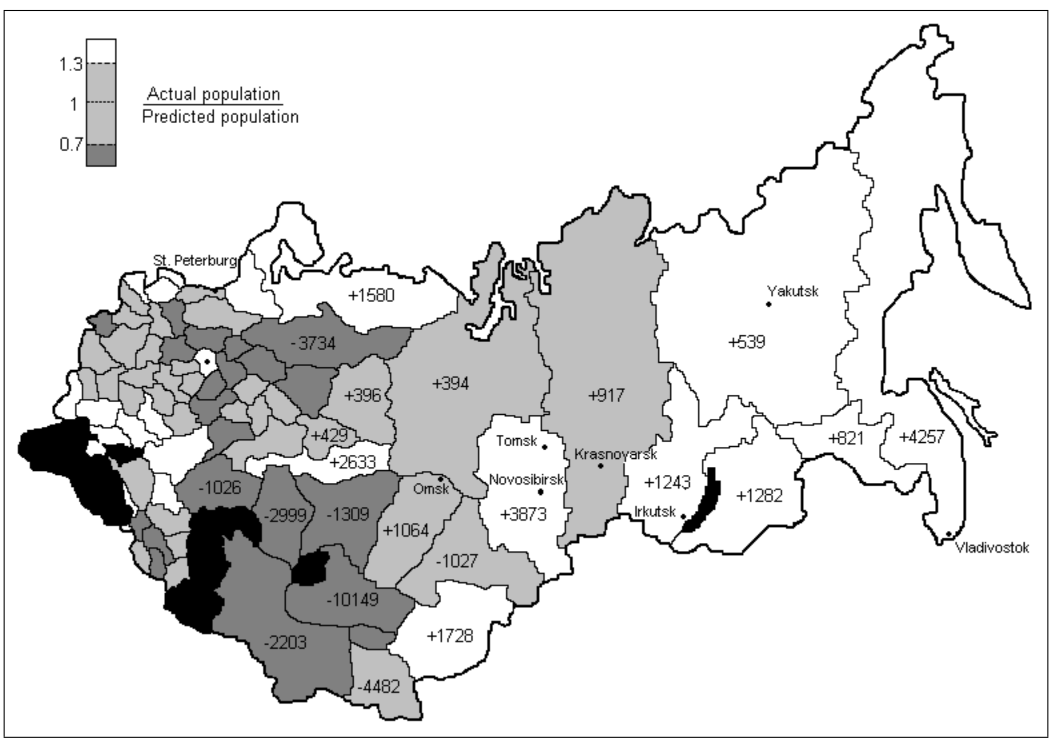


Figure 8: Projected vs. actual population, USSR. Accounted for WWII and corrected for fertility differences (numbers show the absolute difference between projected and actual population, in thousands).

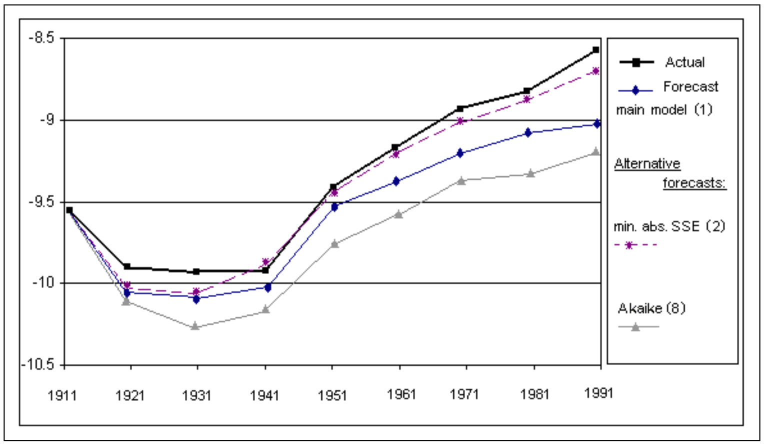
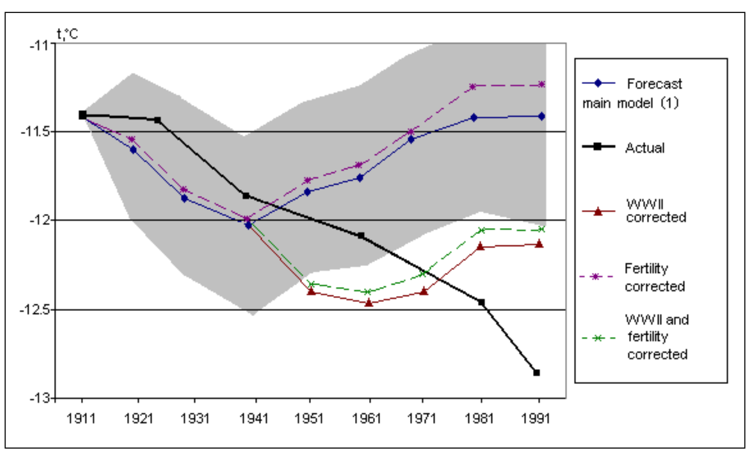


Figure 9: Projected and actual TPC dynamics in Canada. Alternative models compared.



Grey area shows 95% confidence interval around the projected trajectory with no corrections.

Figure 10: Projected and actual TPC dynamics in Russia.