Geography, population density, and per-capita income gaps across US states and Canadian provinces

Nils-Petter Lagerlöf and Syed A. Basher

October 2005

Geography, population density, and per-capita income gaps across US states and Canadian provinces*

Syed A. Basher† and Nils-Petter Lagerlöf‡
Department of Economics, York University, Canada

September 28, 2006

*We are grateful for comments from Tasso Adamopoulos, Ahmet Akyol, David Andolfatto, Berta Esteve-Volart, Paul Klein, Shannon Seitz, and Andrei Semenov. Lagerlöf thanks SSHRC for financial support. We also thank the following persons for help in locating data sources: Katie Davis, Élise Laplaine and Greg LeBlanc of Statistics Canada; Karen D. Thompson of the US Census Bureau; Brian Dahlin and Serah Hyde of the Bureau of Labor Statistics; David Sutherland of NOAA; and Wendell Cox of Demographia.

†E-mail: basher@econ.yorku.ca.
‡Corresponding author. E-mail: lagerlof@econ.yorku.ca.

Abstract: We explain per-capita income gaps across US states and Canadian provinces by the following chain of causation. Geography determined where Europeans originally settled: in Northeastern USA, along those segments of the Atlantic coast where the climate was neither too hot (the US South), nor too cold (Canada). Higher population densities in this early settled region have prevailed to this day. This has in turn affected per-capita incomes because densely populated areas are conducive to skill accumulation; indicatively, many of the world's top universities lie in this region. Our ordinary least-squares regressions show university education having a robust positive and significant effect on per-capita incomes. To control for endogeneity we run various instrumental-variable regressions: some where education today is instrumented with e.g. population density in 1900; and some where different sets of geography variables (e.g. temperature) are used as instruments. Our findings are consistent with the type of causal chain described.

1 Introduction

This paper examines a novel link from geography to the distribution of per-capita incomes across US states and Canadian provinces. The existing literature has often pointed to direct effects of geography, in particular vicinity to waterways and so-called natural harbors, to economic and demographic outcomes. Coastal regions are richer and more populated because they have more trading ports, goes the argument (see, in particular, Rappaport and Sachs 2003).

This explanation fails to account for why the Eastern half of North America is richer and more densely populated than the Western half. It also fails to explain differences in population and per-capita incomes along the Atlantic coast: the US South and Atlantic Canada are relatively poor, and New England is rich.

We believe there is something more involved than just a direct effect from geography. The way we think about these patterns relates to recent work explaining per-capita income gaps across countries through a chain of causation running from geography, via institutions, to current economic outcomes.[1] The econometric approach taken in this literature is to use variables measuring geography and/or demography – such as settler mortality and pre-colonial population density – as exogenous instruments for some measure of institutions, for example an index over protection of property rights.

Using similar econometric techniques we document a related causal chain within the "neo-European" region of the USA and Canada. This chain runs from geography, via early settlement, population density, and education, to economic outcomes today. To see the point, note that five of the the six richest US states – Massachusetts, Connecticut, New York, New Jersey, and Delaware (see Table 1) – lie clustered in a belt along the Atlantic Ocean. This region was the first to be settled and from the start it has had the highest

[1] See Acemoglu et al. (2001, 2002) Easterly and Levine (2003), Glaeser et al. (2004), Rodrick et al. (2004), Acemoglu et al. (2005), Banarjee and Iyer (2005), Lagerlöf (2005), and Tabellini (2005).

population densities in North America. Here lie some of the world's largest cities, like New York City and Boston, and highest ranked universities, such as Harvard, MIT, Yale, Cornell, and Princeton.[2] Surrounding states and provinces have both sparser populations and lower per-capita incomes. (The reader may note that states to the south were all slave states; we return to this below.)

We propose that these patterns are due to geographical fundamentals which made the first Europeans settle in Northeastern USA: along the Atlantic coast, and where the climate was neither too hot (like in the US South), nor too cold (like in Canada). Early settlement has determined population densities and the location of urban centers up to this day, in turn affecting per-capita incomes, because such dense and urban environments are conducive to skill accumulation. This is also why so many top-ranked universities lie in this region.

We combine PPP adjusted per-capita income data across 50 US states and 10 Canadian provinces, with measures of geography (such as average annual temperature, precipitation, and coastal dummies); population densities in 1900 and today; the sex ratio in 1900[3]; urbanization rates today; and the fraction of the population with a university degree today.

We first run a number of ordinary least-squares regressions showing that university education has a robust positive effect on per-capita incomes. It stays significant when controlling for political variables (such as the size of government and unionization); sectoral composition (like fisheries employment); and a Canada dummy.

We then run a number of two-stage least squares regressions where univer-

[2] One could add to this cluster Rhode Island, which is also densely populated and has an Ivy League university (Brown). However, it is not as rich as the other five (see Table 1). Interestingly, in 1840 Rhode Island was the richest state in the union, followed by those other five states; see Easterlin (1960, Table A-1). Why Rhode Island fell behind is a topic left for another paper.

[3] As discussed later, the sex ratio (the number of men per woman) is known in many other contexts to be negatively correlated with urbanization.

sity education is instrumented with historical variables (population density in 1900, the sex ratio in 1900, and slavery in 1850). We find that these historical variables are good instruments for university education, in the sense that they are highly correlated with university education and uncorrelated with the second-stage residual. That is, they seem to affect per-capita incomes through education, rather than directly.

The same conclusion holds when running other instrumental-variable regressions treating population density, the sex ratio, urbanization rates, and university education as endogenous. These are instrumented with various sets of geography variables, such as temperature, rainfall and coastal dummies. Again we find that the instruments are valid: these geography variables seem to affect economic outcomes not directly but rather through their influence on, for example, population density and education.

In short, our results suggest that those five Northeastern states are so rich and densely populated because they lie by the Atlantic coast and have a climate which was inviting to settlers: neither too hot, nor too cold. For the same reason, their too hot, too cold, and too inland neighboring states and provinces are relatively poor and empty today.

We believe the value-added of this exercise is four-fold. First, as argued already, our methodology relates closely to a new empirical development literature on geography, institutions, and income gaps. However, we seem to be the first to think about variation in per-capita incomes within a rich and "Neo-European" region, such as the US and Canada, using a similar instrumental-variable approach as, for example, Acemoglu et al. (2001, 2002).[4]

Second, we propose and test a link from geography to economic outcomes

[4]There is an empirical literature on income gaps and convergence across US states (Barro and Sala-i-Martin 1992, Mitchener and McLean 2003, Berkowitz and Clay 2004). However, none of these use geography variables as instruments for population density and/or education, and do not merge data from US states and Canadian provinces. The only income comparisons across US states and Canadian provinces that we know of is an annual report published by the Fraser Institute (see Karabegović et al. 2004).

which has been largely ignored in the existing literature: demographic and educational variables. This contrasts with e.g. Rappaport and Sachs (2003) who suggest that geography exerts a direct effect on economic outcomes across US regions, for example by affecting trade. It also contrasts with Acemoglu et al. (2001, 2002), and many others, who emphasize institutions (such as property rights) as the intermediate factor between geography and economic outcomes. Within the region we study institutions cannot (other than in a very broad sense) be the only factor involved, since much of the variation in per-capita incomes shows up across institutionally similar states and provinces (like New England and Atlantic Canada). Much of the causality rather seems to run through the rise of urban centers, and the effect these have had on learning and human-capital accumulation.[5] This is also consistent with a vast literature finding that shorter geographical distances facilitate skill accumulation (as discussed in Section 2.3).

Third, we do look at one particular institutional link from geography to economic outcomes: slavery. This institution arose in those regions of the Americas where the climate was suitable to grow staple crops like cotton, tobacco, and sugar: that is, in the Caribbean, Brazil, and the US South.[6] Slavery has had well-documented negative effects on institutions (in the sense of, for instance, voting rights and school reforms), on equality, and on per-capita incomes today. This holds in our data too. More interestingly, however, not only does slavery have a significantly negative effect on per-capita incomes in our regressions; including a slavery variable *strengthens* the effect from population density. In some regressions the effect from population

[5] In a sense, one could argue that our story partly mirrors the growth of Europe's Atlantic regions following the discovery of the Americas (cf Acemoglu et al. 2005). It also relates to Glaeser et al. (2004) who think that European settlers brought not (only) their institutions, but (also) their human capital, although we suggest that population density and urbanization can by itself be a factor behind human capital accumulation and the rise of universities.

[6] See Sokoloff and Engerman (2000), Mariscal and Sokoloff (2000), Engerman and Sokoloff (2002), Acemoglu et al. (2002), and Lagerlöf (2005).

density is significant *only* when we control for slavery. The intuition is that many former slave states are poorer today than their population densities alone would account for. Slavery picks up some of the variation in incomes, thereby also making the population density effect stronger. In that sense, the population-density link and the institutional link seem complimentary.

Fourth, we may add insights to a literature on differences in incomes and other variables across OECD countries, in particular between European countries and the US (Alesina et al. 2001, Gordon 2004, Prescott 2004, Rogerson 2005). Our methodology is a little different: we do not focus on labor supply or taxes; we run regressions rather than calibrating models; and (to iterate) we are the first to look at variation within the US-Canada region. However, our results may be interesting in light of this literature because Canada shares so many characteristics with both Europe and the US. (We return to this discussion in the conclusions in Section 4.)

The rest of this paper is organized as follows. Next Section 2 elaborates on the theory we wish to put forward and discusses how it seems to fit with the data. Section 3 presents the results, first when regressing per-capita incomes on university education and a number of control variables using ordinary least-squares, and then using various instrumental-variable approaches. Section 4 ends with a concluding discussion.

2 The theory and some preliminary evidence

The hypothesis that we are about to investigate can be summarized in a flow chart, as follows:

$$\begin{gathered}\text{Geography} \Rightarrow \text{early settlement}\\ \Rightarrow \text{population density in 1900}\\ \Rightarrow \text{population density today} \Rightarrow \text{accumulation of skills}\\ \Rightarrow \text{per-capita income levels.}\end{gathered}$$

A correlation matrix for some of the variables involved is found in Table 2. Next we discuss the links in more detail.

2.1 Geography and population density in 1900

The first link runs from geography to early settlement. The earliest year for which we have population density data for most states and provinces is around 1900 (only data for Newfoundland is missing). Consider Figure 1 which plots average annual temperature against log population density in 1900. Consistent with Rappaport and Sachs (2003) it can be seen that Atlantic and other coastal states and provinces have higher densities than those located inland, at any given temperature (see Table 1 for a list of codes for all states and provinces). It also seems that the relationship between temperature and population density is ∩-shaped; very hot and very cold states and provinces are less densely populated than those at intermediate temperatures.

This ∩-shaped relationship is even more striking when looking only at the Atlantic region which is relatively homogenous in terms of, for example, mountainousness and rainfall (see Figure 2). As described in the introduction, densities are high in a couple of Northeastern states and lower both to the north and the south of these. Note that there is nothing special about Canada, aside from the weather: in Figure 2 the fitted curve for US states only (dashed) is virtually identical to the fitted curve for states and provinces together (solid). Temperature thus accounts for Canada's sparse population.

Why does temperature have a non-monotonic effect on settlement? It makes sense that early settlers in North America, who were mostly farmers, avoided too cold regions due to their lower agricultural productivity.[7] (The exception may be regions where they could procure food from fishing.) Hot regions may have been unsuitable for European settlers in particular, since they were not resistent to warm-weather diseases (Acemoglu et al. 2001; Coelho and McGuire 1997, 1999). Moreover, much of the migration to the warmer parts of the US were not by free Europeans but by African slaves,

[7] Also in pre-agricultural times humans avoided cold regions probably reflecting availability of food. Anthropologists have noted that density of "cultures" is higher in warmer and wetter regions of the world (Collard and Foley 2002).

without whom population density in the South would have been even lower. As discussed below, geography also seems to have had an effect on incomes through slavery, separate from the population density link.

Other geography variables which seem to have mattered is the number of rainy days per year (Figure 3) and total precipitation (Figure 4). These probably affected agricultural productivity.

All coasts, not only the Atlantic, have higher population densities than inland regions, and so do states and provinces around the Great Lakes. However, the Atlantic is more densely populated than both the Gulf and Pacific coasts (see, for example, Map 2 in Rappaport and Sachs 2003; in fact, the whole eastern half of the US is more densely populated than the western half). The most probable explanation is that the Atlantic is closer to Europe, which is where most early settlers arrived from. Many immigrants stayed in the big cities where their ships landed. Trans-Atlantic trade may also have had an impact on early growth of the Atlantic region of North America, as it did on the European side of the Atlantic (Acemoglu et al. 2005).

2.2 Population density in 1900 and today

Despite migration to, from, and within North America log population density in 1900 is highly correlated with log population density in 2001 (Table 2 and Figure 5). One may note that places like Arizona and Nevada have had relatively fast population growth over the last century, probably due to inventions like air-conditioning (Rappaport 2004). Others, like Prince Edward Island, have seen population numbers fall. But here we do not want to focus on why some regions have grown and others contracted. Rather we want to emphasize what really stands out in Figure 5: that those states and provinces which were the most densely populated in 1900 are so still today.

2.3 Population density and education

The idea that a shorter geographical distance between people enhances the exchange of ideas and accumulation of skills goes back at least to Jacobs (1969), and probably much longer (Glaeser 1999 quotes Alfred Marshall on agglomeration effects). Empirical support can be found in, for example, Jaffe et al. (1993), who show that patent citations are negatively related to distance. Glaeser and Maré (2001) find that wages are higher in cities because cities promote learning rather than the reverse causality by which skilled people choose to live in cities. Theoretical foundations can be found in, for example, Glaeser (1999).

It is also possible that colleges and universities, due to scale effects in education, have come to be located in regions with dense populations, both historically and today. Many Ivy League universities lie in the densely populated region around Northeastern USA, and vicinity to educational institutions seems to matter for educational choice: Card (1995) finds that men who grew up near a four-year college have higher education and earnings, also when controlling for regional factors and family background; Glaeser and Saiz (2003) find that cities of a given size grow faster if they have more colleges per capita. A skilled labor force can also attract high-technology industries (Henderson et al. 1995).

2.3.1 Urbanization

Obviously, population per unit of land area over a whole state or province may not be the best measure of the mechanism we try to capture. Ideally one would want a measure of how well “connected” people are to the type of social networks which build skills and/or enhance growth of high-skilled industries. Alternatively one may want data over how far the average resident of a state or province is from the closest university or college.

Lacking any such data we look at urbanization rates for 2000. This is positively correlated with log population density in 2001 (see Table 2 and Figure 6). The fit is not terribly good, however, possibly because the urbanization measure we use is the fraction of the population living in cities exceeding the modest size of 1,000 people (we use this measure because it is the most comparable between Canada and the US). The fraction living in larger cities could perhaps have provided a better fit.[8] Moreover, urbanization rates may not be exactly the right measure either: population density may just as well serve as a proxy for whatever is the "true" measure.

These issues aside, we note that the signs are right: both log population density in 2001 and our measure of urbanization rates are positively correlated with the fraction of the population having a university degree (urbanization slightly less than population density; see Table 2 and Figures 7 and 8).

2.3.2 The sex ratio in 1900

We do not have historical urbanization data (at least not for both Canada and the US) but a good proxy could be the sex ratio, that is, the number of men per woman. Edlund (2005) documents that rural areas in the Western world are relatively short on women, compared to urban areas. This seems particularly true in new settlements in colonial times. Guttentag and Secord (1983, Ch. 5) document that in frontier societies of the US men vastly outnumbered women into the 20th century, while the situation was rather the opposite in New England. (See also Angrist 2002.) This fits with our data, where the sex ratio in 1900 is strongly negatively correlated with log population density (Figure 11).

2.4 Education and per-capita incomes

The fraction with a university degree is highly positively correlated with per-capita incomes, notably more so than are population density in 2001

[8] We did look at data over the population sizes of different cities, in order to estimate the number of people living in cities with, for example, 100,000 people or more. Given that we look at state and province data, however, one problem with that approach is that many cities belong to more than one state or province.

and urbanization (see Table 2). This suggests that the link from population density and urbanization to per-capita incomes does work through human capital.

2.5 Slavery

As seen in Figures 14 and 15 the fraction slaves in the population in 1850 varies with geographical variables, such as temperature and rainfall, which facilitated the growth of staple crops (see the discussion in the introduction). Slavery also seems to impact per-capita incomes and education today. The link is there when considering all states and provinces (Figure 16), and (more visibly) among slave states (Figures 17 and 18). Slavery thus constitutes a link from geography to economic outcomes, which does not work through population density.

3 Regression results

3.1 Ordinary least-squares regressions

Table 3 presents the ordinary least-squares results when regressing log per-capita income (GSP or GPP) on the fraction having a university degree and a number of other variables. University education has a high explanatory power on its own [an R-squared of 52.3% in column (1)]. It also stays very significant when controlling for a range of other control variables.

In all specifications but the first we enter a Canada dummy, which is mostly insignificant. One may note that it is consistently positive: when controlling for levels of university education (and a number of other variables) Canada does not seem to be poorer than the US, but rather richer. However, the Canada dummy tends to have a negative sign in the second stage of some of the instrumental-variable regressions shown later.

Some political variables are significant. Consider first the variables from the Fraser Institute, indicated by FI in Table 3. Among these, the size of government, discriminatory taxation, and union density come out as significant (union density barely). These variables, however, are hard to interpret. They are all 1-10 indices, and it is not even always clear what a high or low score means (see Section A.3.6 in the appendix). Moreover, even though the sources do describe some of the details about how the FI variables are computed, the raw data is not provided and we have not been able to replicate these indices.

There is another problem with the FI variables. We computed an alternative set of political variables, with a clearer interpretation: the ratio of federal, state, and local government expenditure to incomes. As seen in column (10) of Table 3, a high ratio of federal expenditure to income has a significantly negative effect on per-capita income. But the direction of causality is far from obvious. Notably, this variable is the highest and lowest for two Canadian provinces: Prince Edward Island (which is the poorest of all 60 states and provinces), and Alberta (the richest Canadian province). This probably reflects the Canadian federal government's choices in response to existing income gaps, rather than exogenous causes behind the gaps. In other words, Ottawa does not *make* Albertans rich by not giving them money; they *are* rich and thus do not get any money.

The state and local spending ratios, on the other hand, are more plausible causes of per-capita income differences. However, as seen in column (11) and (12) these are insignificant, and of the "wrong" sign, respectively (that is, a bigger local government is associated with higher per-capita incomes).

Moreover, out of the three ratios the federal one is the most strongly correlated with FI's size-of-government variable (the correlation coefficient is -0.83). This is not strange because the FI variable is based on similar data. However, it does suggest that the size of government as measured by the Fraser Institute shows up as significant in these regressions, not because it causes income gaps, but rather because it is caused by existing gaps through federal expenditure.

Another set of variables measures the sectoral composition of the state's or province's labor force. The employment share in fisheries has a negative, but statistically insignificant, effect on per-capita incomes [column (14)]; the employment share in natural resources industries (mostly oil and gas) has a positive and significant effect [column (15)]. See also Figures 12 and 13.

However, data over fishery employment is available for only 33 states and provinces, and natural resource employment for 55 states and provinces. Most of the variation is among a few states and provinces, like Alaska, Alberta, and Atlantic Canada (cf Figures 12 and 13). Endogeneity is also an issue. Canada's Atlantic provinces may have come to rely more on fishery today because when fishery began its decline the labor force did not move into other sectors. Whatever prevented the growth of non-fishery industries should be the ultimate cause of current income gaps; we believe that cause is population density and education.

To sum up, the results shown in Table 3 suggest that university education has a robust positive and significant relationship with per-capita income levels. However, it is not clear whether education causes these income gaps, or if the causality goes the other way around. To address that issue we next turn to instrumental-variable analysis.

3.2 Two-stage least squares regressions

3.2.1 Historical variables as instruments

Table 4 shows the results from some two-stage least squares regressions where university education is instrumented for using three historical variables. These are: log population density in 1900, the sex ratio (the number of men per woman) in 1900, and the fraction slaves in the population in 1850.

Each of these variables is theoretically a good candidate for an instrument. We have already argued that population density can exert a causal impact on human-capital accumulation, and that the sex ratio may serve as a useful proxy for urbanization. (One shortcoming with the sex ratio is that data is missing for 8 states and provinces.) Slavery seems to have had a negative impact on education (cf Figure 18). The ultimate reason should be that slaves were often forbidden (or otherwise prevented) from learning to read or write. Also after abolition these effects seem to have lingered on. School reforms have tended to come later in formerly slave-dependent Caribbean countries and Brazil compared to Canada and the US north of the Chesapeake Bay (Mariscal and Sokoloff 2002), and there are indications of similar patterns within the US South (Lagerlöf 2005).

Using measures from 1850 and 1900, rather than today, could also help alleviate some reverse-causality concerns: for example, high levels of education (and income) today may spur migration to cities today, but population density a century back seems more likely to causally impact education today, rather than the other way around. (However, there are some caveats to this reasoning, as discussed below.)

The two-stage least squares results in Table 4 show that per-capita income is affected positively by the fraction with a university degree. This holds when this fraction is instrumented with the slavery variable, together with either log population density [column (1)], or the sex ratio [column (2)], as well as when using all three variables as instruments [column (3)].

To be valid instruments these variables should first of all be highly correlated with the instrumented variable. As seen in the lower panel of Table 4, the first-stage regressions do not have a very high R-squared (about 15%). However, an F-test shows that the instruments are jointly significant. Moreover, although the first-stage estimated coefficients on log population density and the sex ratio are insignificant, their signs are the expected ones. That is, log population density has a positive effect on education, whereas slavery and the sex ratio have negative effects. However, the sex ratio gets the wrong sign when entered jointly with log population density since these two variables are highly correlated.

For the instruments to be good they should also not influence the dependent variable (log per-capita income) other than through the instrumented variable (the fraction with a university degree). In other words, the instruments should be uncorrelated with the error terms in the second-stage regression. This seems to be the case: Hansen's J-test does not reject the hypothesis that the second-stage error terms are uncorrelated with the instruments; that is, the p-value is very high (well above conventional risk levels of 5% or 10%). To see that the instruments do not exert any direct impact on per-capita incomes we also report the results when letting each instrument enter the second-stage regression [columns (4) to (6)]. As seen, they all come out as insignificant.

All in all, it seems that these historical variables are valid instruments. However, in order to truly believe that they are valid we must also believe that the direction of causality runs from the instruments (population density, the sex ratio, and slavery) to the instrumented variable (the fraction with university degree). The fact that the instruments date a century, or longer, back in time is no guarantee that this is the case. For example, some third factor may have made some states and provinces rich and highly educated, and thus induced migrants to settle there, and thereby also made the economy less dependent on slave labor. In principle, what regions become prosperous and densely populated may be due to chance, and work through the coordination of many agents' simultaneous decisions. However, we have already argued that there is one fundamental determinant of settlement in North America: geography.

3.2.2 Geography as instruments

Our story builds around a chain of causation. Geography mattered for where Europeans originally settled which has come to determine population densities today which has impacted per-capita income patterns because densely populated regions (cities) are conducive to skill accumulation. To test this hypothesis using an instrumental-variable approach we must make two choices. First, we must decide what set of geography variables (temperature etc.) to use as instruments. Second, we must choose what intermediate variables (that is, what links in the chain: population density, education, etc.) to be instrumented. Generally, the results depend on how these choices are made: some instruments do a better job together with some endogenous (instrumented) variables; other combinations do not work equally well.

A smaller set of geography instruments Table 5 shows the results from a couple of two-stage least squares regressions where as instruments we use annual temperature and its square, and an Atlantic dummy (cf Figures 1 and 2). These are used to instrument log population density in 1900 or 2001, or the sex ratio in 1900. The dependent variable is log per-capita income.

As seen from the odd-numbered columns [(1), (3), and (5)], the results seem discouraging at first: for none of the three instrumented variables is the coefficient in the second-stage regression significant at any conventional risk level.

However, we recall from Figures 14 to 18 and the results in Table 4 that geography affects education and per-capita income also through slavery. We thus allow the fraction slaves in 1850 to enter the regression as an endogenous variable [columns (2), (4), and (6)]. As seen, not only does slavery come out as significant – the other three instrumented variables do too (at least at the 10% level), and they are also larger in size.

This result is quite interesting. We may think of slavery as an institutional variable capturing a source of per-capita income gaps similar to the historic forces at play in other regions of the Americas. As we discussed earlier, such mechanisms are emphasized by, for example, Sokoloff and Engerman (2000), Acemoglu et al. (2001, 2002), Engerman and Sokoloff (2002), and Lagerlöf (2005). They argue that geographical fundamentals like soil and climate lead to the introduction of slave-based plantation production, which lead to post-abolition inequality, and growth retarding institutions. Here we focus on another chain of causation: from geography to the locations of early settlements, population density up to this day, and skill accumulation in such densely populated areas. This is somewhat related to an argument made by Glaeser et al. (2004): that European settlers may have brought their human capital rather than institutions. So, in our data, does geography affect per-capita incomes through population density or through institutions? The answer is: both. It is not a matter of either/or; in fact, the effect from population density is seen only when controlling for institutions (that is, slavery).[9]

To understand why consider the plot of per-capita incomes and log population density in 2001 in Figure 9. As seen, those states which had the highest fraction slaves – Mississippi, South Carolina, and Louisiana – are outliers. They are thus poorer than their population densities can account for. When entering slavery into the regressions it picks up some of that variation in incomes, in effect making the population density effect stronger.

It is also interesting to note from the even-numbered columns in Table 5 that the Canada dummy comes out as significant in the first-stage regression but not in the second-stage regression when the instrumented variable is log population density in 2001; and vice versa when the instrumented variable is historical: either log population density in 1900 or the sex ratio in 1900. That is, there was nothing special about Canada by 1900: the sex ratios and population densities of Canadian provinces can be explained by geography. However, there is a negative "Canada effect" working after 1900 both on contemporary per-capita incomes and population densities.

The even-numbered columns of Table 5 also show that the instruments are good, in the sense that they perform well by Hansen's overidentification test: zero correlation between the instruments and the second-stage residuals cannot be rejected based on the J-test, with p-values around 23-28%. The instruments are also jointly significant in the first-stage regression, as seen from the F-test in the lower panel (although the temperature variables do a poor job for the sex ratio).

[9] As we shall see soon, it seems that even though population density and slavery are two different links between geography and economic outcomes, both work through education.

A larger set of geography instruments Not only temperature and vicinity to the Atlantic have had an impact on population density. In Table 6 we use a larger set of geography variables as instruments. These are: average temperature and its square, average precipitation, average number of rainy days, and coastal dummies: for the Atlantic only in odd-numbered columns; and for the Atlantic and the Great Lakes in even-numbered.

We now use as endogenous (instrumented) variables log population density in 1900 and 2001, as well as two variables through which these have supposedly affected per-capita incomes: urbanization rates, and university education. Slavery is also treated as endogenous and instrumented with the same variables (we do not report the first-stage regression).

Again, as seen in Table 6, we find that the instrumented variables overall have a significant effect on per-capita incomes, most with p-values below 5%. When we instrument urbanization rates the effect is insignificant when using the Atlantic dummy only [column (5)], but becomes significant at the 10% level when adding the Great Lakes dummy [column (6)].

The instruments seem to be valid too. The F-test in the first-stage regression suggests the instruments are jointly significant (even though many of the instruments are individually highly insignificant). Also, Hansen's J-test for overidentification suggests that the instruments are uncorrelated with the residuals in the second-stage regression.

We may also note that slavery comes out as insignificant when university education is instrumented but not when population density is instrumented. This suggests that the negative effect from slavery to economic outcomes works through education. This is consistent with, for example, Mariscal and Sokoloff (2000) and Lagerlöf (2005).

4 Discussion and concluding remarks

In 1999, out of 60 US states and Canadian provinces Connecticut was the richest with a per-capita Gross State Product of US$46,245. Prince Edward Island was the poorest with a per-capita Gross Province Product of US$20,545. Their per-capita income ratio was thus about 2.25. This paper is about explaining such income gaps.

One may argue that this is not an interesting research topic: 2.25 is not a huge gap by international standards; across countries per-capita income gaps can exceed a factor of 30 or 40. One reason that we still find this topic interesting is that different North American regions share a lot (though not everything) in geography, history, institutions, language, and ethnicity. Comparing income gaps across states and provinces of North America thus amounts to keeping such factors constant, to some extent, and this can teach us a great deal also about cross-country income gaps.

Our explanation builds around a chain of causation. Geography mattered for where Europeans settled which determined where cities and urban centers are located today. This has impacted per-capita income patterns because cities are conducive to skill accumulation. We test this hypothesis with various instrumental variable specifications and the data does not reject it.

Population density may play some role for cross-country income gaps too: it usually shows up with a positive sign in per-capita income regressions (Olson 1996), and city-states like Luxembourg, Hong Kong, and Singapore are rich. However, it seems plausible that in a world-wide context geography may exert a much stronger effect through institutions; in the region we study the population-density effect shows up more clearly because the US and Canada have so similar institutions.

Some would suggest other explanations than those we propose. Within the US-Canada region the poorest locations lie in Atlantic Canada (Table 1). It may thus be tempting to attribute income gaps across this region to variations in the dependence on fishery. But this is not an exhaustive explanation, we argue. There are other poor regions which do not rely on fishery (West Virginia is landlocked and almost as poor as Nova Scotia); Alaska is rich and has a relatively large fraction of its employment in fishery. Fishery data is only available for 33 states and provinces, only a few of these have any significant fishery industries, and our regression results are not too supportive of a fishery explanation (see Section 3 and Figure 13).

Moreover, it is not obvious why the decline of one sector of the economy would necessarily mean the decline of a whole region. If fishery has declined, why have people previously employed in the fisheries not moved to other sectors? Consider Massachusetts, a maritime region which once had a fishery and whaling industry, and is rich today and not dependent on fishery. What makes Nova Scotia different from Massachusetts? Why could not Halifax be like Boston? A good explanation should be deeper than simply pointing to the decline of some sector of the economy. It should point to fundamental causes (like geography), rather than proximate.

Our results are also interesting when thinking about income gaps between European countries, US states, and Canadian provinces. The fact that the richest regions of the United States are the most densely populated suggests that per-capita income gaps between Europe and the US cannot be explained by the same factor, since most European countries are poorer and more densely populated than the US. Some would explain Europe's relative poverty by emphasizing that the US has a smaller government, lower taxes, and a less regulated labor market. In the US-Canada case such explanations have been put forward by the Fraser Institute. We do not rule such explanations out but here they do not seem to tell the whole story. For example, we find that while government expenditure (relative to income) is negatively correlated with per-capita income, the correlation holds only for federal expenditures but not for state and local. The causation thus seems to go from poverty to spending, rather than the other way around. That is, poor regions receive more transfers from benevolent politicians in Ottawa and Washington.

One may also note that those five rich Northeastern states listed in the introduction – Massachusetts, Connecticut, New York, New Jersey, and Delaware – all voted for John Kerry in 2004. In many ways they seem to resemble Europe (or Canada) in their political preferences. If conservative policies made these states rich they were hardly chosen by the states' own electorates. The differences in policies between the US and Europe (and Canada) may rather originate in differences in political institutions, as suggested by Alesina et al. (2001): the political and electoral system in the US makes it harder for socialist parties to establish themselves. One could even hypothesize that maybe those rich Northeastern states would have been poorer if they had been able to elect their own federal administrations and imposed European-style labor laws and tax codes. Ironically, it may be voters in poorer "red" states inland who ensure the prosperity of more left-leaning voters in Massachusetts by keeping conservative and low-tax politicians in the White House. This hypothesis relates to how some on the political left in the US see things: see, for example, Thomas Frank's (2005) "What's the Matter with Kansas?"

A Data appendix

Below we list our data sources, some of which are available online only. When it is not self-explanatory we try to describe in as much detail as possible what steps to take to access the online data. All data are also available as a STATA file at:

http://www.arts.yorku.ca/econ/lagerloef/HP/PubDataUSCan.dta

A.1 Geography variables

A.1.1 Temperature, precipitation, and rainy days

Data for three weather variables were retrieved from the Weatherbase Web site at www.weatherbase.com. These are: average temperature over the year (in degrees Fahrenheit); average precipitation (inches of rainfall over the year); and average number of rainy days per year.

Where available these refer to the capital of the state or province (see Table 1 for a list). Else another city was chosen alphabetically. For temperature and precipitation, data from Nova Scotia refer to Ecum Secum; all other temperature and precipitation data refer to the capital.

Rainy days data for capitals were often missing, in which case we used data for these cities: Banff, Alberta; Abbotsford, British Columbia; Brandon, Manitoba; Belle Isle, Newfoundland; Ecum Secum, Nova Scotia; Armstrong, Ontario; Alma, Quebec; Bowling Green, Kentucky; Aberdeen, Maryland; Alexandria, Minnesota; Belton, Missouri; Battle Mountain, Nevada; Atlantic City, New Jersey. For all other states and provinces we used the capital.

A.1.2 Coastal dummies

The following states and provinces are considered to be located on the Atlantic coast: New Brunswick, Newfoundland, Nova Scotia, Prince Edward Island, Quebec, Connecticut, Delaware, Florida, Georgia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, North Carolina, Rhode Island, South Carolina, and Virginia.[10]

The following states and provinces are considered to be located by the Great Lakes: Ontario, Illinois, Indiana, Michigan, Minnesota, New York, Ohio, Pennsylvania, and Wisconsin.

A.2 Historical variables

A.2.1 The fraction slaves in 1850

The fraction slaves in the population is calculated as the total number of slaves in 1850 over total population in 1850. This data is made available by the Geospatial and Statistical Data Center at the University of Virginia Library. Their Web site is at:

http://fisher.lib.virginia.edu/collections/stats/histcensus/

Canada did not use slavery.

A.2.2 The sex ratio in 1900

The numbers of males and females in Canada refer to the year 1901 and are from Census of Canada (1902, Table III).

The corresponding data for the US were extracted from the Geospatial and Statistical Data Center, University of Virginia Library (see the previous section; click on 1900 and follow the links).

Note that the Canadian data is for 1901 and the US data is for 1900, but we refer to this variable as the sex ratio in 1900.

[10]Depending on whether the Delaware Bay belongs to the Atlantic, or not, we could categorize Pennsylvania as Atlantic; we decided not to, but few results would change if we did.

A.2.3 Population density in 1900

The Canadian population density data are from Series A54-66: Population density per square mile, Canada and provinces, 1871 to 1976, Statistics Canada (1983).

US population density is from Table 5, Statistical abstract of the United States 1901, which is available at

http://www.census.gov/statab/www/

Note that the Canadian data is for 1901 and the US data is for 1900, but we refer to this variable as population density in 1900.

A.3 Contemporary variables

A.3.1 GDP per capita

PPP adjusted GDP per capita for both Canadian provinces and US states are from the Web site Demographia run by the Wendell Cox Consultancy, at www.demographia.com. All figures are for 1999 and in current US dollars. The exact links are:

http://www.demographia.com/db-cangdpr99.htm (for Canada)
http://www.demographia.com/db-usgdpr99.htm (for the US)

Note that the state and province level equivalents of GDP are called GSP (Gross State Product) and GPP (Gross Province Product), respectively. In the text we also call this variable per-capita income for short.

A.3.2 Fishery and natural resource data

Natural resource employment By employment in natural resource industries we mean, for Canada, employment in mining and oil and gas extraction. These numbers are from Statistics Canada (2002, Table A.32). The numbers for the US are people employed in natural resource and mining industries, from the Bureau of Labor Statistics, available online at:

http://data.bls.gov/PDQ/outside.jsp?survey=sm

For both Canada and the US the numbers are from 2000.

Fishery employment The number of people employed in fisheries in Canada are downloaded electronically from the Statistics Canada Web site, at www.statcan.ca. To access the data, select census; select data; select topic-based tabulations; click on number 11 – "Canada's workforce: paid work." Table 8 provides the number of people employed by industry and province.

The US data are from Pritchard (2003, p. 95), under the category "employment, craft, and plants." The US data refer to total employment in both the fish-processing and wholesale industry.

For both Canada and the US these numbers are from 2001.

Total employment To get fishery and natural resource employment as fractions of total employment we use the following data. For Canadian provinces, 2000 and 2001 total employment is from Statistics Canada (2003, Table 18), available online at www.statcan.ca.

For the US total employment in 2000 is from Table 572, Statistical Abstract of the United States 2001, US Census Bureau, available online at http://www.census.gov/statab/www/. Total employment for 2001 is from Table 565, Statistical Abstract of the United States 2002, also available online at http://www.census.gov/statab/www/.

A.3.3 Urbanization rates

Our measure of the urbanization rate is the fraction of the population living in an urban area with 1,000 persons or more. For Canadian provinces the data were downloaded electronically from the Statistics Canada Web site, www.statcan.ca, following these steps: select census; select data; select population and dwelling counts; select "urban and rural."

For the US the corresponding numbers are from the US Census Bureau's 2000 Summary File 1 (SF 1) 100-Percent data (Table P2), available online at http://factfinder.census.gov.

For both Canada and the US these numbers are from 2000.

A.3.4 Population density in 2001

Population density per square kilometer for Canadian provinces was downloaded electronically from the Statistics Canada Web site, at www.statcan.ca, using the following steps: select learning resources; select E-STAT; select table of contents; select people; select data; select population and demography; from the census databases select population characteristics; select 2001 population and dwelling counts; select the desired item from the list. One square mile is 2.59 square kilometers.

For US states population density per square mile is collected from Table 19, Statistical abstract of the United States 2002, US Census Bureau, available online at http://www.census.gov/statab/www/.

For both Canada and the US these numbers are from 2001.

A.3.5 Fraction with a university degree

For Canada this fraction is given by the number of persons 15 years and over with a university (Bachelor) degree, divided by total population 15 and over. The number of persons with a degree are from Statistics Canada, 2001 Census, downloaded electronically from the Statistics Canada Web site, www.statcan.ca, following these steps: select census; select search by topic; select education in Canada: school attendance and levels of schooling; click on number 1 (under topic-based tabulations) – detailed highest level of schooling. Total population numbers are from the same Web site: select census; select search by topic; select age and sex; click on number 2, "profile of age and sex."

The US data is the fraction of the population with a Bachelor degree or more for persons 25 years old and over, based on the 2000 census, from

Table 231, Statistical Abstract of the United States 2003, US Census Bureau, available at http://www.census.gov/statab/www/.

The Canadian data is from 2001 and the US data from 2000.

A.3.6 Political variables

Fraser Institute indicators We use six political indicators constructed by the Fraser Institute (FI), a Canadian think tank. These are meant to measure "economic freedom" and/or the flexibility of labor markets and take values on a scale from 1 to 10. Indicators A to D below are from Karabegović et al. (2004); indicators E and F from Clemens et al. (2004).

Indicator A: "Size of the government" is an index measuring general government consumption expenditures relative to GSP or GPP. A higher score means smaller government.

Indicator B: "Discriminatory taxation" is an index measuring how "discriminatory" the tax system is; taxation is considered discriminatory if, for instance, the link between taxes paid and services received is weak, or marginal taxes are high. A higher score means less discriminatory taxation.

Indicator C: "Minimum wage legislation" is an index measuring the annual income earned by someone working at the minimum wage relative to per-capita GSP or GPP. A higher score means a higher minimum wage.

Indicator D: "Union density" is an index measuring the fraction of the work force who is unionized. A higher score means that a larger fraction is unionized.

Indicator E: "Average duration of unemployment" is an index measuring just that. A higher score means longer unemployment duration.

Indicator F: "Flexibility in labor-relation laws" is an index measuring the flexibility in different areas of labor law. A higher score means that labor laws are more flexible.

Government expenditure Aside from the variables from the Fraser Institute we calculated three other political variables: the ratio of federal, state/provincial, and local government expenditure to income.

Data over government expenditures across Canadian provinces are from Statistics Canada (2003, Table 7-9). We first divided expenditure by population to get it in per-capita terms; we then divided by per-capita personal income to get the total expenditure as a fraction of income. Both population and per-capita personal income by province were collected from Statistics Canada (2003, Table 18).

The same data for US states are from Sagoo (2005; Tables C18, E15, and F14). To get total expenditure as a fraction of income we divided per-capita expenditures by per-capita personal income (collected from Table A14 in Ibid.).

These sources are partly the same as those used to calculate FI's index over the size of government (Indicator A above).

For both Canada and the US the numbers are from 2002.

References

[1] Angrist, J., 2002, How do sex ratios affect marriage and labor markets? Evidence from America's second generation, Quarterly Journal of Economics 117, 997-1038.

[2] Acemoglu, D., S. Johnson, and J.A. Robinson, 2001, The colonial origins of comparative development: an empirical investigation, American Economic Review 91, 1369-1402.

[3] − − −, 2002, Reversal of fortune: geography and institutions in the making of the modern world income distribution, Quarterly Journal of Economics 117, 1231-1294.

[4] − − −, 2005, The rise of Europe: Atlantic trade, institutional change and growth, American Economic Review, 95, 546-579.

[5] Alesina, A., E.G. Glaeser, and B. Sacerdote, 2001, Why doesn't the US have a European-style welfare state?, mimeo, Harvard University.

[6] Banerjee, A., and L. Iyer, 2005, History, institutions, and economic performance: the legacy of colonial land tenure systems in India, American Economic Review 95, 1190-1213.

[7] Barro, R.J., and X. Sala-i-Martin, 1992, Convergence, Journal of Political Economy 100, 223-251.

[8] Berkowitz, D. and K. Clay, 2004, Initial conditions, institutional dynamics and economic performance: evidence from American states, mimeo, University of Pittsburgh.

[9] Card, D., 1995, Using geographic variation in college proximity to estimate the return to schooling, in: Christofodes, L.N., E.K. Grant, R. Swidinsky (eds.), Aspects of labour market behaviour – essays in honour of John Vanderkamp, University of Toronto Press, Toronto.

[10] Census of Canada, 1902, Fourth census, volume 1: population, printed by S.E. Dawson, Ottawa.

[11] Clemens, J., K. Godin, A. Karabegović, and N. Veldhuis, 2004, Measuring labour markets in Canada and the United States, Fraser Institute.

[12] Coelho, P.R.P., and R.A. McGuire, 1997, African and European bound labor in the New World: The biological consequences of economic choices, Journal of Economic History 57, 83-115.

[13] − − −, 1999, Biology, diseases, and economics: an epidemiological history of slavery in the American South. Journal of Bioeconomics 1, 151-190.

[14] Collard, I.F., and R.A. Foley, 2002, Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules?, Evolutionary Ecology Research 4, 371-383.

[15] Easterlin, R.A., 1960, Interregional differences in per capita income, population, and total income, 1840-1950, in: Trends in the American economy in the nineteenth century, Princeton University Press (for the NBER).

[16] Easterly, W. and R. Levine, 2003, Tropics, germs, and crops: the role of endowments in economic development, Journal of Monetary Economics, 50, 3-39.

[17] Edlund, L., 2005, Sex and the city, Scandinavian Journal of Economics 107, 25-44.

[18] Engerman, S.L., and K.L, Sokoloff, 2002, Factor endowments, inequality, and paths of development among New World economies, Economia 3, 41-109.

[19] Frank, T., 2005, What's the matter with Kansas? How conservatives won the heart of America, Owl Books, New York.

[20] Glaeser, E.L., 1999, Learning in cities, Journal of Urban Economics 46, 254-277.

[21] Glaeser, E.L., and D.C. Maré, 2001, Cities and skills, Journal of Labor Economics 19, 316-342.

[22] Glaeser, E.L., and A. Saiz, 2003, The rise and fall of the skilled city, mimeo, Harvard University.

[23] Glaeser, E.L., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer, 2004, Do institutions cause growth?, Journal of Economic Growth 9, 271-303.

[24] Gordon, R., 2004, Two centuries of economic growth: Europe chasing the American frontier. mimeo, Northwestern University.

[25] Guttentag, M., and P.F. Secord, 1983, Too many women? The sex ratio question, Sage Publications, Beverly Hills, California.

[26] Henderson, V., A. Kuncoro, M. Turner, 1995, Industrial development in cities, Journal of Political Economy 103, 1067-1090.

[27] Jacobs, J., 1969, The economy of cities, Random House, New York.

[28] Jaffe, A.B., M. Trajtenberg, R. Henderson, 1993, Geographic localization of knowledge spillovers as evidenced by patent citations, Quarterly Journal of Economics 108, 577-598.

[29] Karabegović, A., F. McMahon, D. Samida, and G. Mitchell, 2004, Economic freedom of North America, annual report, Fraser Institute.

[30] Lagerlöf, N.P., 2005, Geography, institutions, and growth: the United States as a microcosm, mimeo, York University.

[31] Mariscal, E., and K.L. Sokoloff, 2000, Schooling, suffrage, and the persistence of inequality in the Americas, 1800-1945, in: Haber, S. (ed.) Political institutions and economic growth in Latin America – essays in

policy, history, and political economy, Hoover Institution Press, Stanford, California.

[32] Mitchener, K.J., and I.W. McLean, 2003, The productivity of US states since 1880, Journal of Economic Growth 8, 73-114.

[33] Olson, M., 1996, Big bills left on the sidewalk: why some countries are rich, and others poor, Journal of Economic Perspectives 10, 3-24.

[34] Prescott, E.C., 2004, Why do Americans work so much more than Europeans?. Federal Reserve Bank of Minneapolis Quarterly Review 28, 2-13.

[35] Pritchard, E.S. (ed.), 2003, Fisheries of the United States 2002, National Marine Fisheries Services, NOAA, Silver Spring, Maryland, http://www.st.nmfs.gov/st1/fus/current/2002-fus.pdf

[36] Rappaport, J., 2004, Moving to nice weather, RWP 03-07 Federal Reserve Bank of Kansas City.

[37] Rappaport, J., and J.D. Sachs, 2003, The United States as a coastal nation, Journal of Economic Growth 8, 5-46.

[38] Rodrik, D., A. Subramanian, and F. Trebbi, 2004, Institutions rule: the primacy of institutions over geography and integration in economic development, Journal of Economic Growth 9, 131-165.

[39] Rogerson, R., 2005, Structural transformation and the deterioration of European labor market outcomes, mimeo, Arizona State University.

[40] Sachs, J.D., 2001, Tropical underdevelopment, NBER working paper 8119.

[41] Sagoo, S. (ed.), 2005, Facts and figures on government finance, 38th Edition, The Tax Foundation, Washington, D.C. http://www.taxfoundation.org/news/show/147.htm

[42] − − −, 2003, Institutions don't rule: direct effects of geography on per-capita income, NBER working paper 9490.

[43] Statistics Canada, 1983, Electronic edition of historical statistics of Canada, second edition, Catalogue no. 11-516-XIE, http://www.statcan.ca/english/freepub/11-516-XIE/sectiona/toc.htm

[44] − − −, 2002, Human activity and the environment, Annual statistics, Catalogue No. 16-201-XIE, http://estat.statcan.ca/content/english/articles/other/other-envi1.pdf

[45] −−−, 2003, Provincial Economic Accounts, Catalogue No. 13-213-PPB, http://www.statcan.ca:8096/bsolc/english/bsolc?catno=13-213-P

[46] Sokoloff, K.L., and S.L. Engerman, 2000, History lessons: Institutions, factor endowments, and paths of development in the New World, Journal of Economic Perspectives 14, 217-232.

[47] Tabellini, G., 2005, Culture and institutions: economic development in the regions of Europe, mimeo, IGIER, Bocconi University.

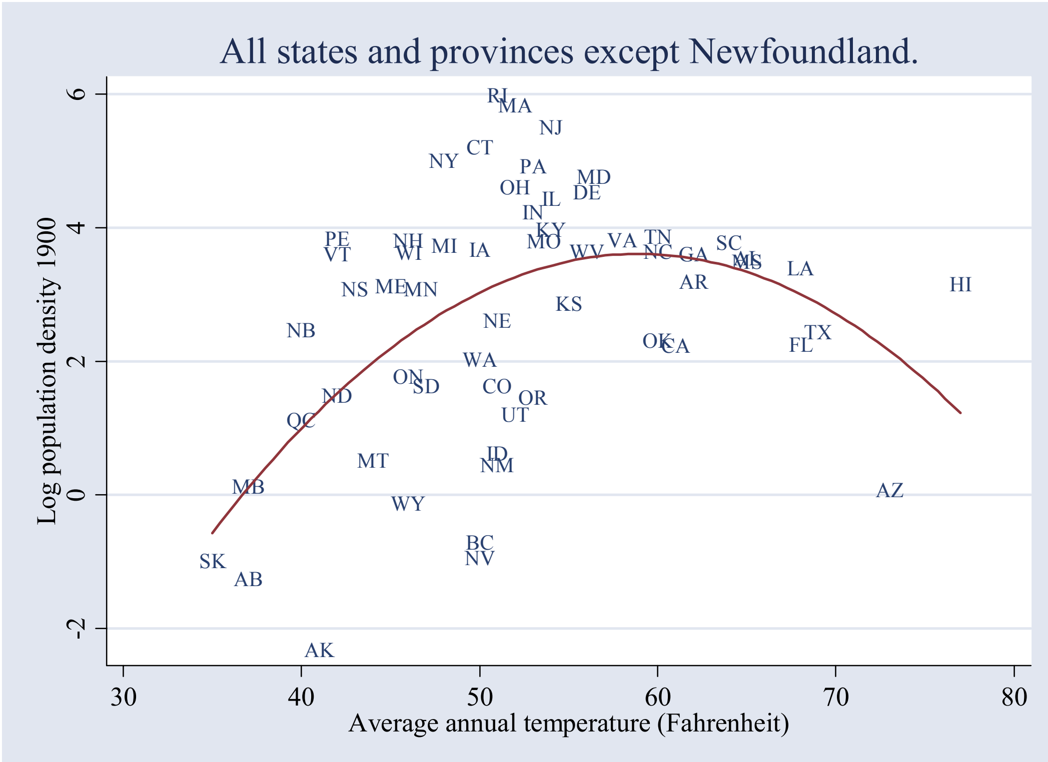

Figure 1. Temperature and population density.

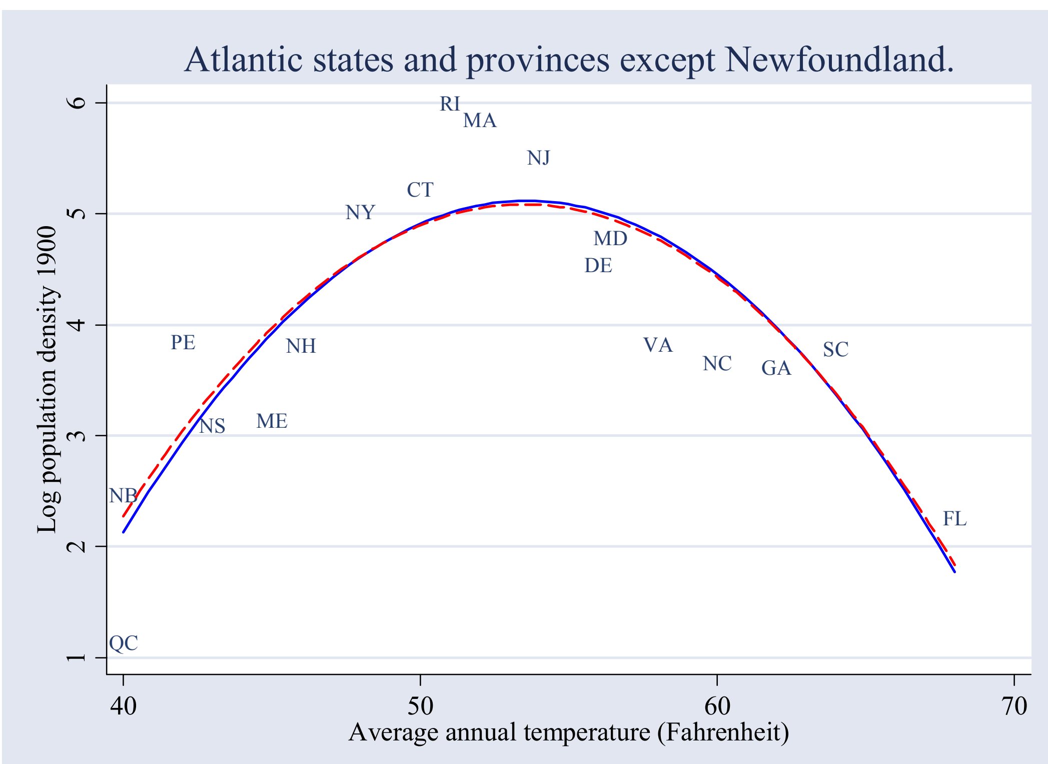

Figure 2. Temperature and population density for Atlantic states and provinces. Fitted curves for states and provinces (solid) and states only (dashed).

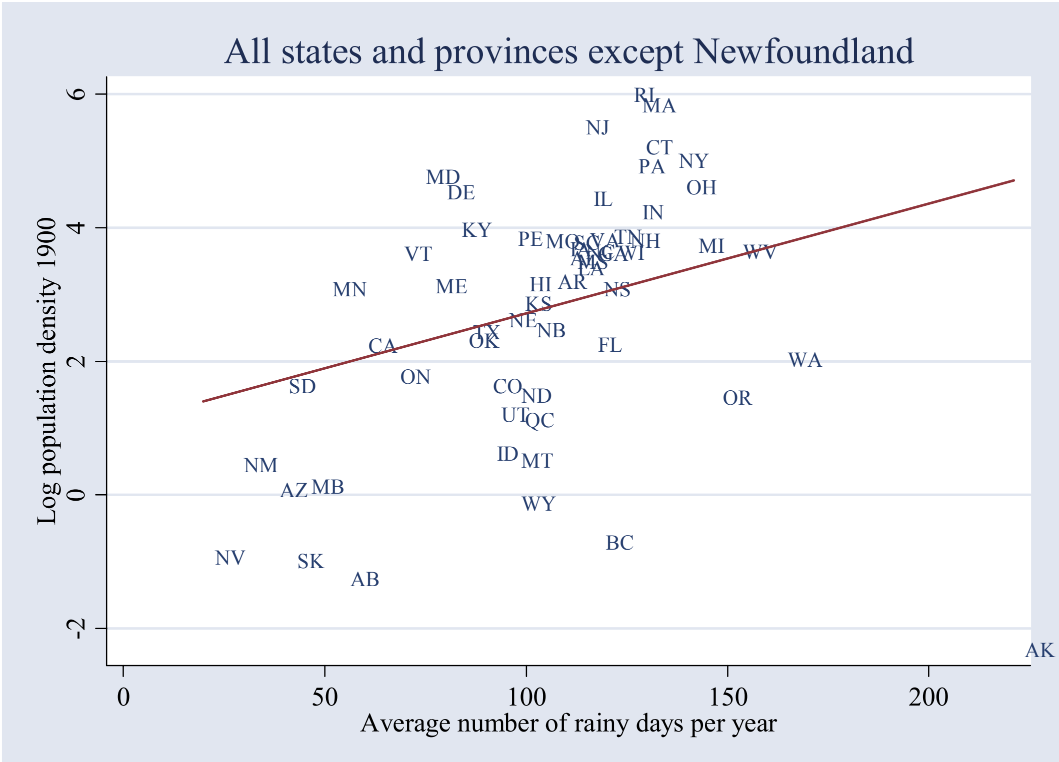

Figure 3: Population density and number of rainy days.

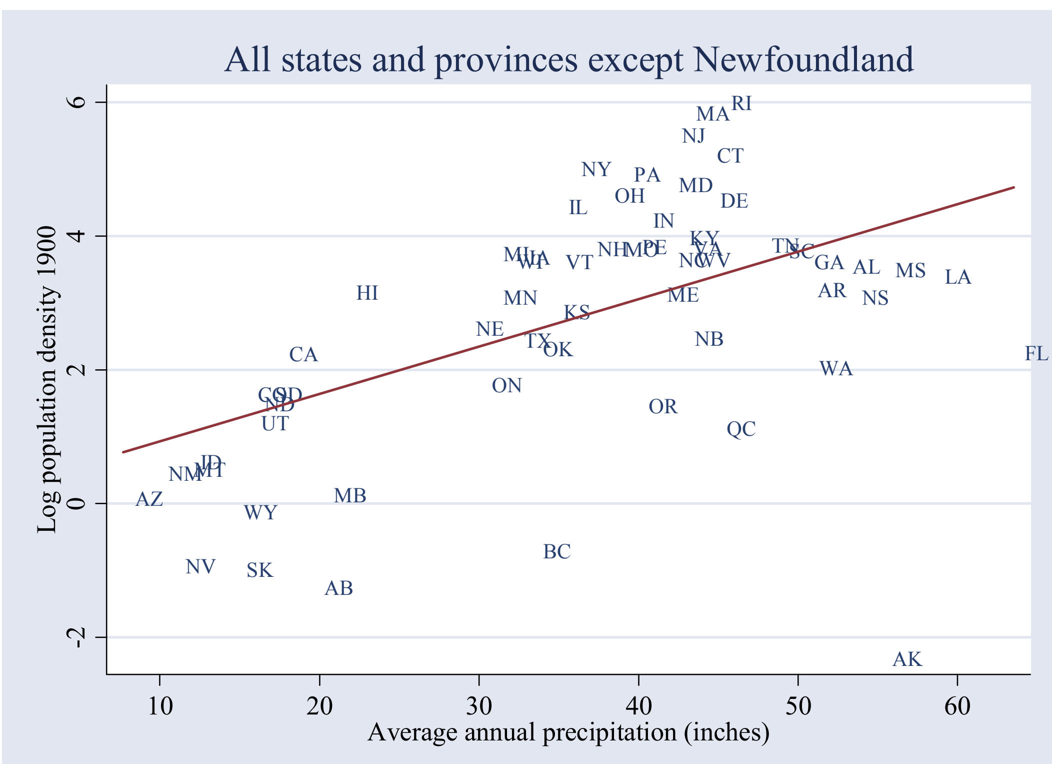

Figure 4: Population density and total rainfall.

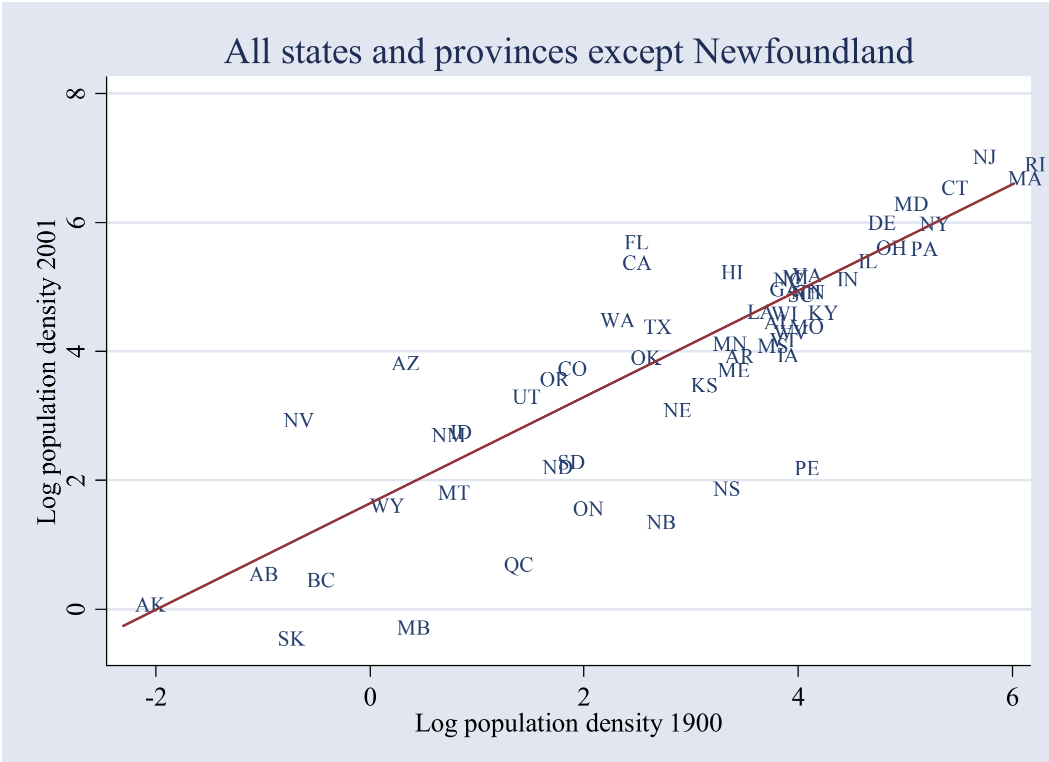

Figure 5: Population density a century ago and today.

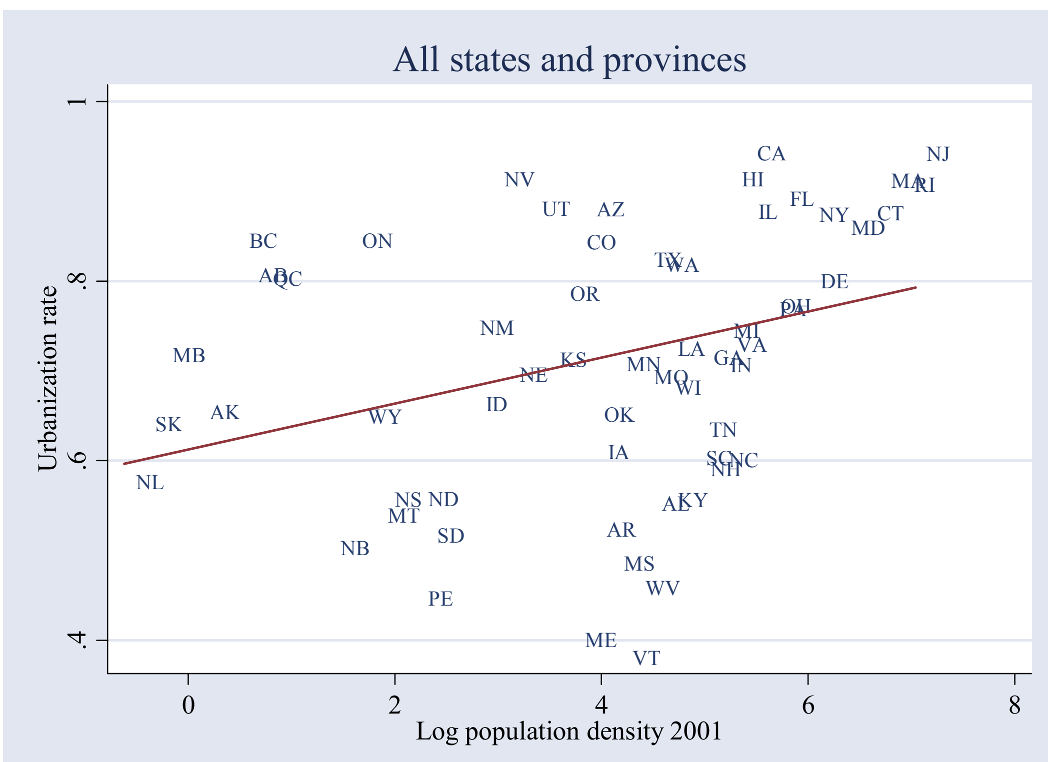

Figure 6: Population density and urbanization.

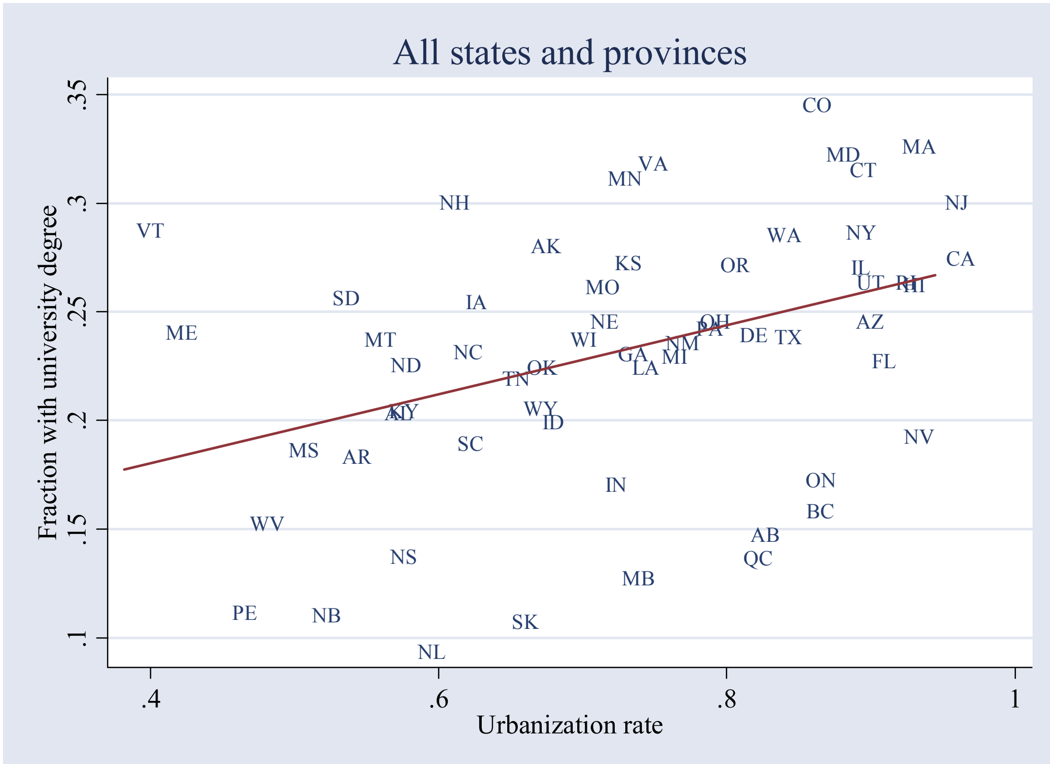

Figure 7: Urbanization and university education.

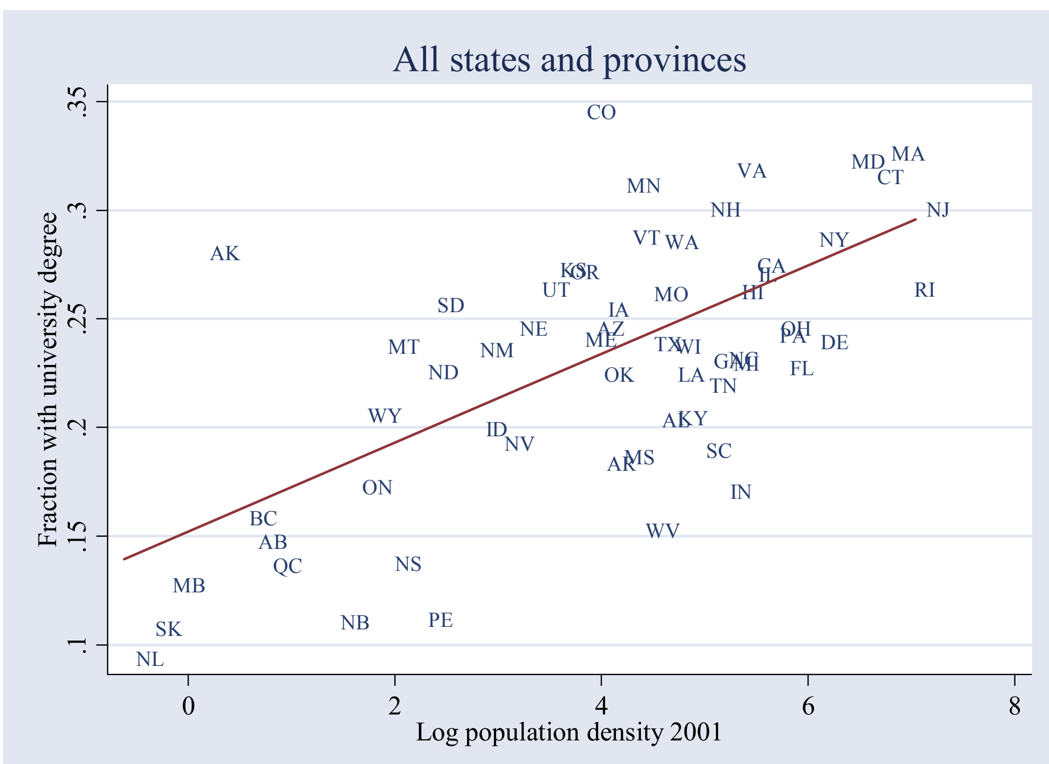

Figure 8: Population density and university education.

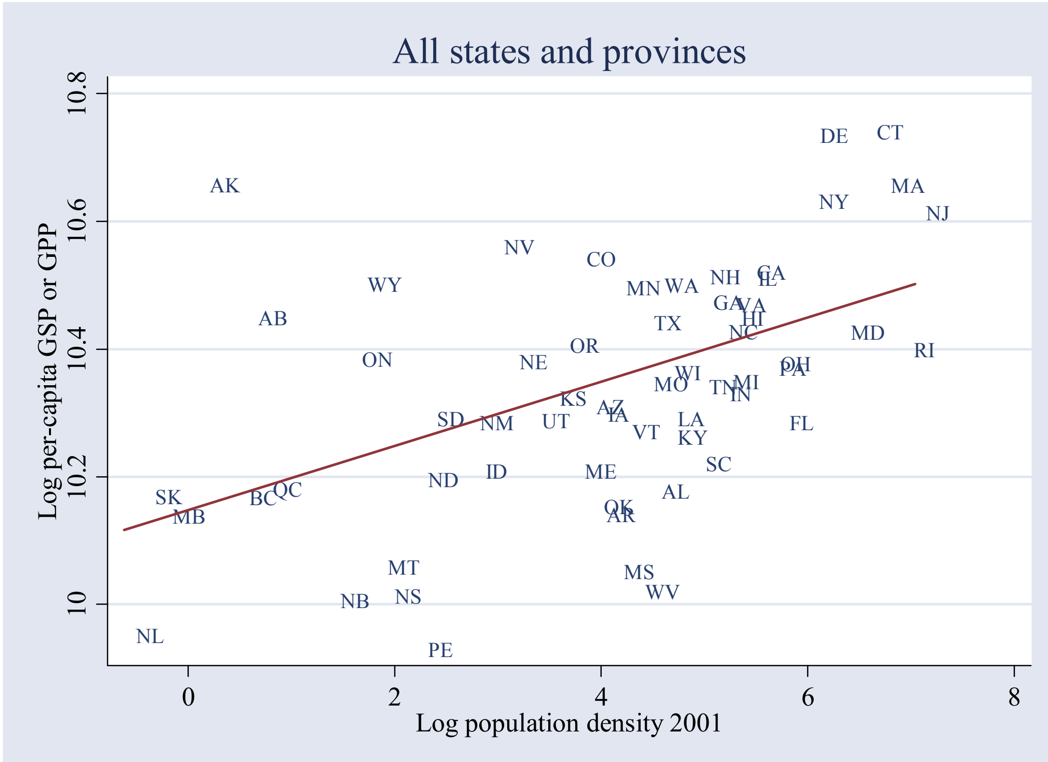

Figure 9: Population density and per-capita income.

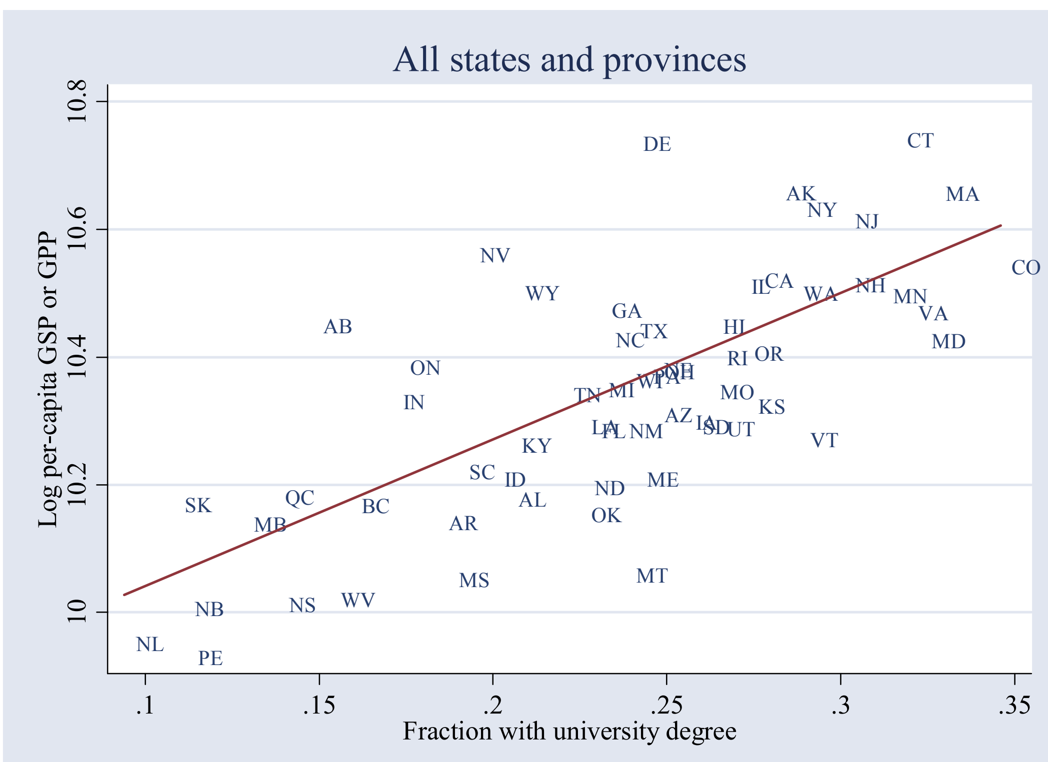

Figure 10: University education and per-capita income.

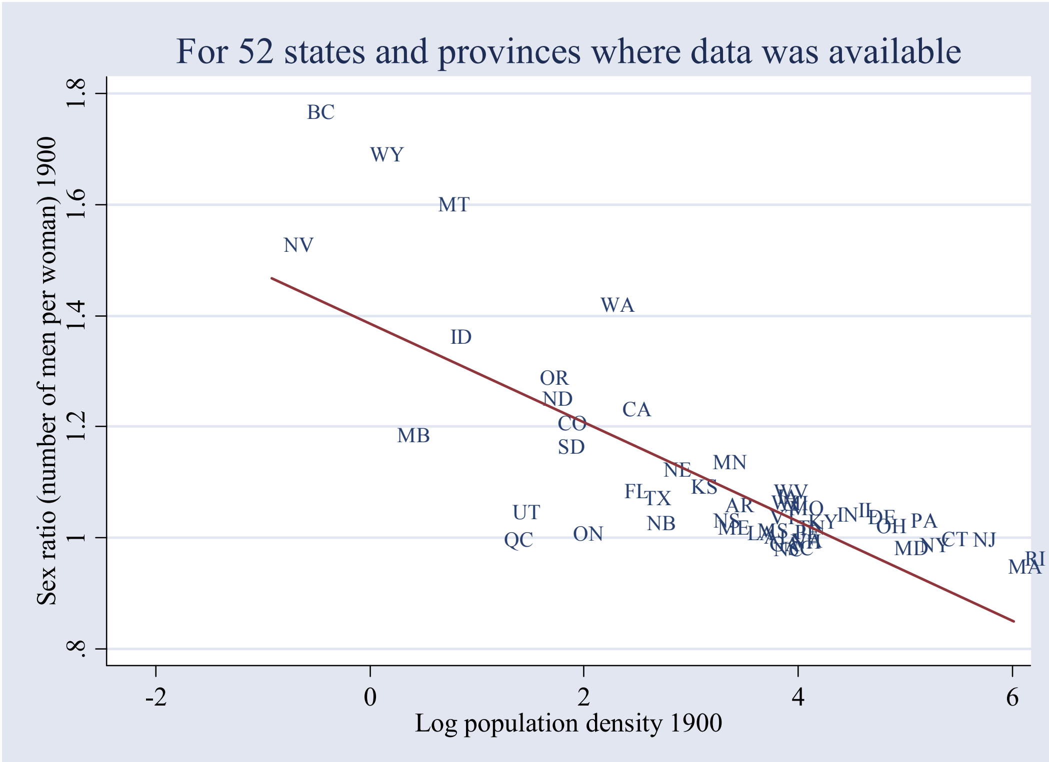

Figure 11: Sex ratio and population density a century ago.

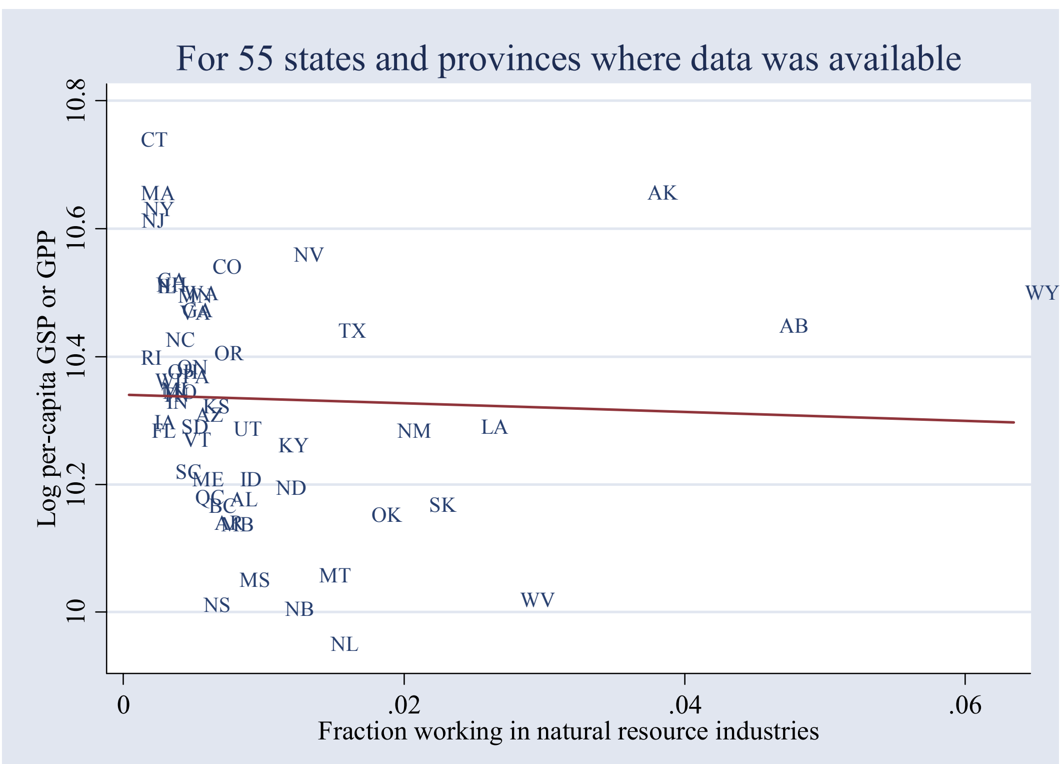

Figure 12: Employment in natural resource industries (e.g. oil, gas, and mining) and per-capita income.

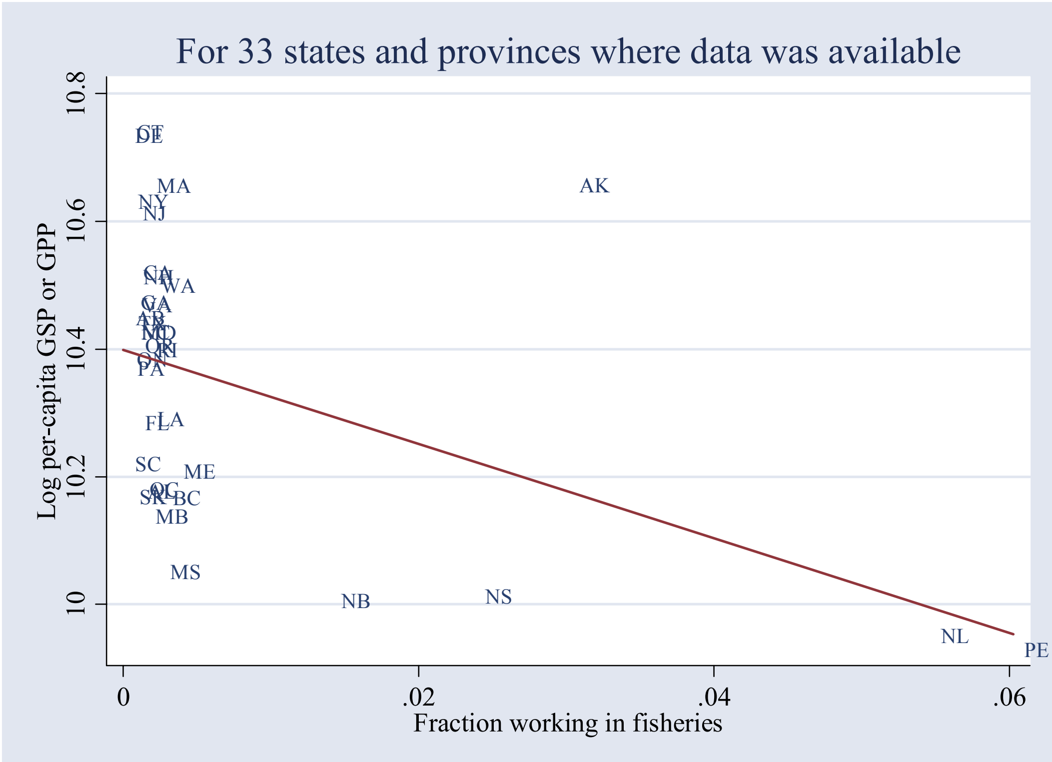

Figure 13: Employment in fisheries and per-capita income.

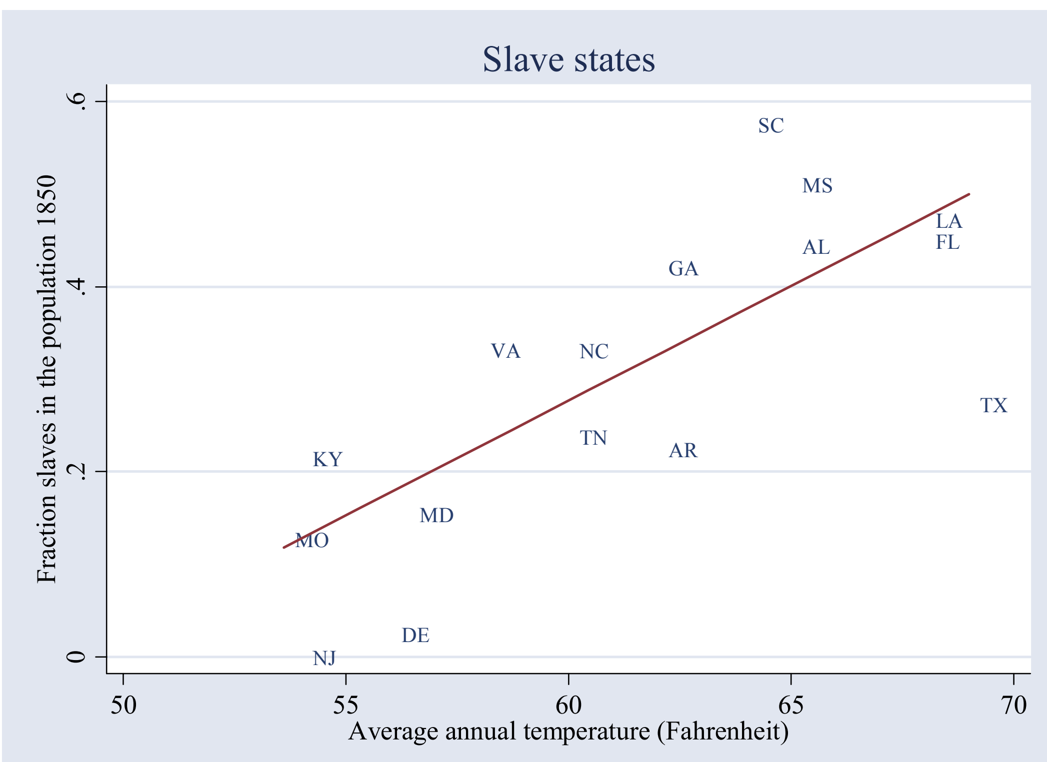

Figure 14: Slavery and temperature across slave states.

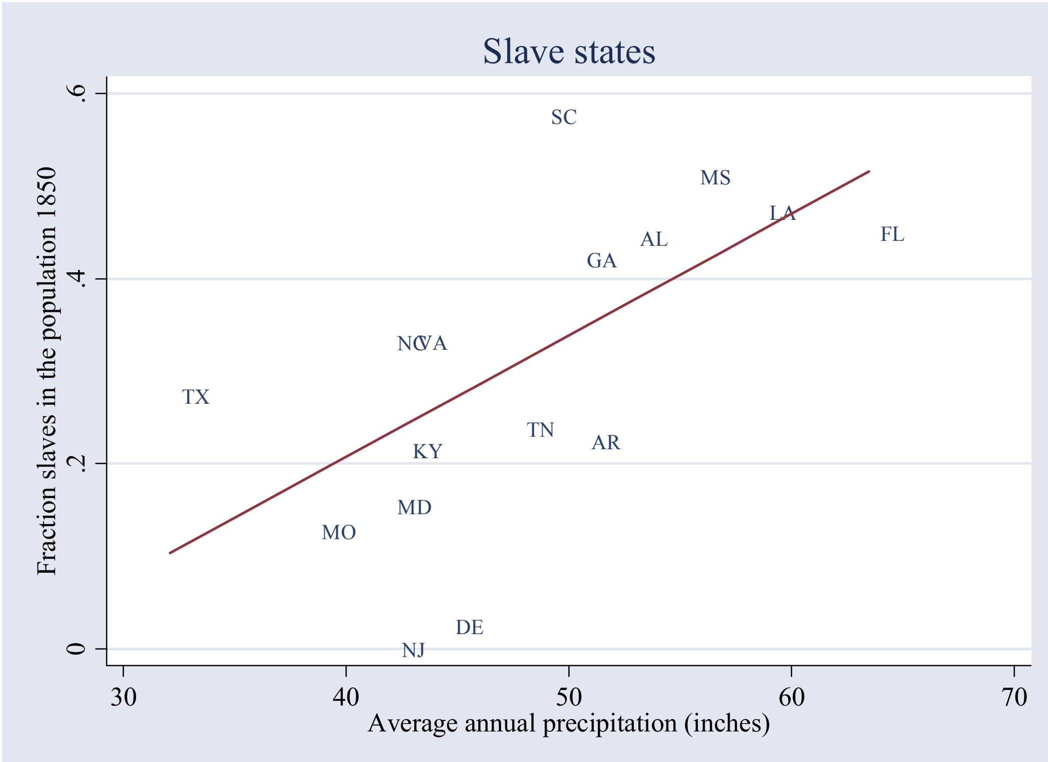

Figure 15: Slavery and rainfall across slave states.

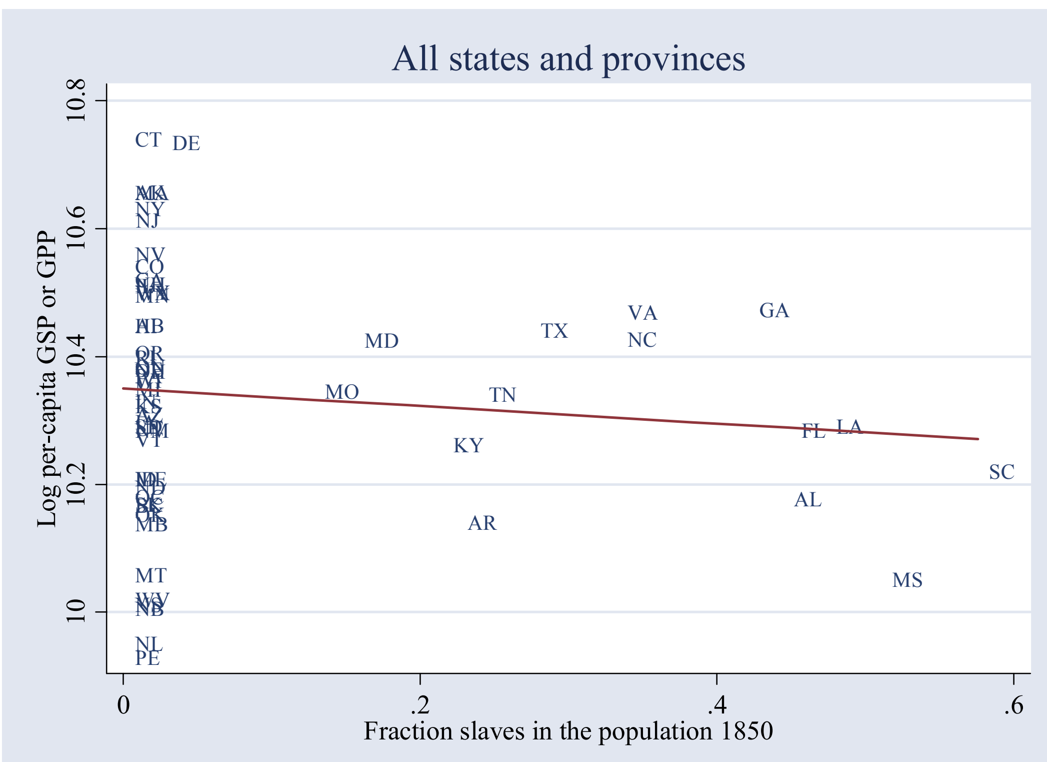

Figure 16: Slavery and per-capita income across all states and provinces.

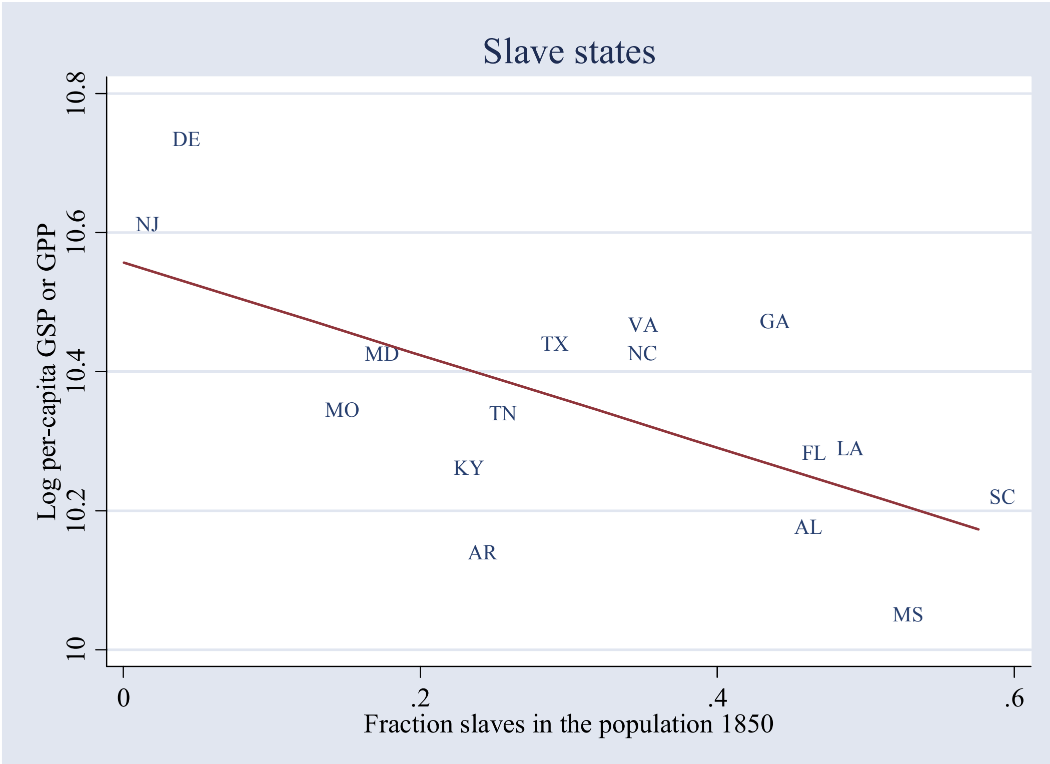

Figure 17: Slavery and per-capita income across slave states.

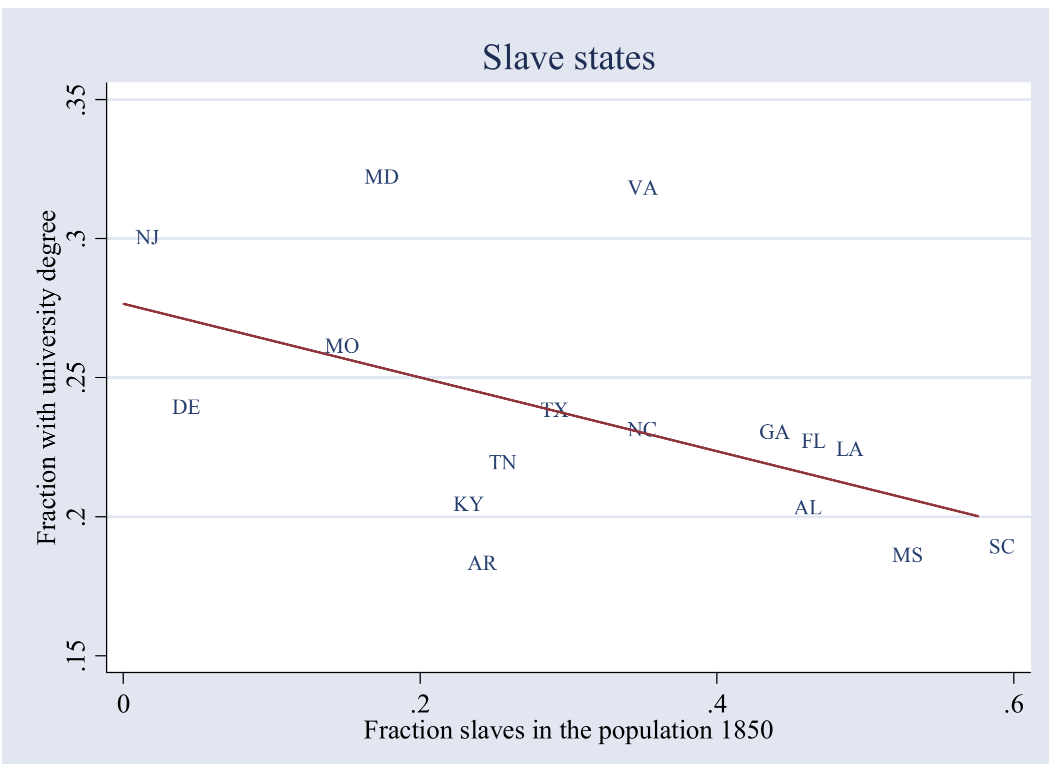

Figure 18: Slavery and university education across slave states.

Table 1: List of states/provinces and per-capita incomes

Rank	State or province	Code	Capital	Per-capita GSP or or GPP (US$), 1999	Rank	State or province	Code	Capital	Per-capita GSP or or GPP (US$), 1999
1	Connecticut	CT	Hartford	46,245	31	Tennessee	TN	Nashville	31,017
2	Delaware	DE	Dover	46,008	32	Indiana	IN	Indianapolis	30,659
3	Alaska	AK	Juneau	42,539	33	Kansas	KS	Topeka	30,460
4	Massachusetts	MA	Boston	42,519	34	Arizona	AZ	Phoenix	30,070
5	New York	NY	Albany	41,469	35	Iowa	IA	Des Moines	29,707
6	New Jersey	NJ	Trenton	40,713	36	South Dakota	SD	Pierre	29,505
7	Nevada	NV	Carson City	38,615	37	Louisiana	LA	Baton Rouge	29,496
8	Colorado	CO	Denver	37,900	38	Utah	UT	Salt Lake City	29,411
9	California	CA	Sacramento	37,082	39	New Mexico	NM	Santa Fe	29,328
10	New Hampshire	NH	Concord	36,823	40	Florida	FL	Tallahassee	29,309
11	Illinois	IL	Springfield	36,746	41	Vermont	VT	Montpelier	28,908
12	Wyoming	WY	Cheyenne	36,380	42	Kentucky	KY	Frankfort	28,665
13	Washington	WA	Olympia	36,352	43	South Carolina	SC	Columbia	27,515
14	Minnesota	MN	Saint Paul	36,223	44	Maine	ME	Augusta	27,185
15	Georgia	GA	Atlanta	35,402	45	Idaho	ID	Boise	27,183
16	Virginia	VA	Richmond	35,243	46	North Dakota	ND	Bismarck	26,814
17	**Alberta**	**AB**	**Edmonton**	**34,540**	47	**Quebec**	**QC**	**Quebec City**	**26,432**
18	Hawaii	HI	Honolulu	34,512	48	Alabama	AL	Montgomery	26,333
19	Texas	TX	Austin	34,288	49	**Saskatchewan**	**SK**	**Regina**	**26,094**
20	North Carolina	NC	Raleigh	33,799	50	**British Columbia**	**BC**	**Victoria**	**26,086**
21	Maryland	MD	Annapolis	33,782	51	Oklahoma	OK	Oklahoma City	25,724
22	Oregon	OR	Salem	33,079	52	Arkansas	AR	Little Rock	25,388
23	Rhode Island	RI	Providence	32,848	53	**Manitoba**	**MB**	**Winnipeg**	**25,328**
24	**Ontario**	**ON**	**Toronto**	**32,373**	54	Montana	MT	Helena	23,376
25	Nebraska	NE	Lincoln	32,259	55	Mississippi	MS	Jackson	23,220
26	Ohio	OH	Columbus	32,157	56	West Virginia	WV	Charleston	22,516
27	Pennsylvania	PA	Harrisburg	31,931	57	**Nova Scotia**	**NS**	**Halifax**	**22,336**
28	Wisconsin	WI	Madison	31,708	58	**New Brunswick**	**NB**	**Fredericton**	**22,187**
29	Michigan	MI	Lansing	31,257	59	**Newfoundland**	**NL**	**St. John's**	**21,008**
30	Missouri	MO	Jefferson City	31,174	60	**Prince Edward Island**	**PE**	**Charlottetown**	**20,545**

Notes: Canadian provinces in bold. GSP stands for Gross State Product, GPP for Gross Province Product. Incomes are PPP adjusted.

Table 2: Correlation matrix

	Log per-capita GSP or GPP	Fraction with university degree	Urbanization rate	Log population density 2001	Log population density 1900
Log per-capita GSP or GPP	1.000				
Fraction with university degree	0.723	1.000			
Urbanization rate	0.647	0.393	1.000		
Log population density 2001	0.508	0.654	0.334	1.000	
Log population density 1900	0.176	0.361	0.009	0.841	1.000

Notes: Correlation coefficients based on all 60 states and provinces, except for the last row which excludes Newfoundland (which lacks population density data for 1900).

Table 3: Higher education and per-capita income: ordinary least-squares regressions

	Dependent variable is log per-capita GSP or GPP							
	Panel A: specifications (1) to (8)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	9.811 (0.000)	9.719 (0.000)	9.231 (0.000)	9.206 (0.000)	9.684 (0.000)	9.926 (0.000)	9.566 (0.000)	9.596 (0.000)
Fraction with university degree	**2.297 (0.000)**	**2.642 (0.000)**	**1.661 (0.000)**	**2.311 (0.000)**	**2.627 (0.000)**	**2.456 (0.000)**	**2.716 (0.000)**	**2.709 (0.000)**
Canada dummy		0.074 (0.280)	0.128 (0.016)	0.207 (0.004)	0.055 (0.457)	0.012 (0.862)	0.164 (0.252)	0.112 (0.129)
Size of government (FI)			0.100 (0.000)					
Discriminatory taxation (FI)				0.104 (0.000)				
Average unem-ployment duration (FI)					0.003 (0.483)			
Union density (FI)						-0.022 (0.059)		
Felixibility in labor laws (FI)							0.015 (0.474)	
Minimum-wage legislation (FI)								0.016 (0.172)
No. of observations	60	60	60	60	60	60	60	60
R-squared	0.523	0.533	0.746	0.643	0.537	0.562	0.537	0.548

	Panel B: specifications (9) to (16)							
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Constant	9.184 (0.000)	10.184 (0.000)	9.747 (0.000)	9.540 (0.000)	10.045 (0.000)	9.615 (0.000)	9.595 (0.000)	9.396 (0.000)
Fraction with university degree	**1.347 (0.000)**	**1.898 (0.000)**	**2.573 (0.000)**	**2.698 (0.000)**	**1.997 (0.000)**	**3.193 (0.000)**	**3.015 (0.000)**	**3.894 (0.000)**
Canada dummy	0.032 (0.748)	0.032 (0.580)	0.062 (0.376)	0.147 (0.060)	0.094 (0.180)	0.145 (0.125)	0.124 (0.055)	0.210 (0.004)
Size of government (FI)	0.084 (0.000)							
Discriminatory taxation (FI)	0.069 (0.006)							
Average unem-ployment duration (FI)	0.006 (0.073)							
Union density (FI)	-0.019 (0.038)							
Felixibility in labor laws (FI)	-0.018 (0.392)							
Minimum-wage legislation (FI)	0.008 (0.329)							
Ratio of federal expenditure to income		-1.174 (0.000)			-1.212 (0.000)			
Ratio of state expenditure to income			-0.021 (0.378)		0.026 (0.240)			
Ratio of local expenditure to income				1.416 (0.060)	0.944 (0.160)			
Fraction working in fisheries						-2.479 (0.156)		-2.530 (0.115)
Fraction working in na-tural resource industries							3.313 (0.027)	5.188 (0.004)
No. of observations	60	60	60	60	60	33	55	30
R-squared	0.825	0.677	0.539	0.562	0.693	0.717	0.595	0.850

Notes: FI denotes data from the Fraser Institute; see Section A.3.6 in the appendix for explanations of what a high/low score means*P* -values in parentheses.

Table 4: Two-stage least squares regressions: history as instruments

	Dependent variable is log per-capita GSP or GPP Instrumented variable is fraction with university degree					
	Second-stage results					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	9.686 (0.000)	9.619 (0.000)	9.565 (0.000)	9.556 (0.000)	9.467 (0.000)	9.624 (0.000)
Fraction with university degree	**2.776** **(0.000)**	**3.046** **(0.000)**	**3.265** **(0.001)**	**3.314** **(0.000)**	**3.411** **(0.000)**	**3.038** **(0.252)**
Canada dummy	0.100 (0.393)	0.081 (0.483)	0.106 (0.359)	0.109 (0.260)	0.120 (0.230)	0.077 (0.819)
Log population density 1900				-0.001 (0.955)		
Sex ratio 1900					0.055 (0.712)	
Fraction slaves in 1850						-0.029 (0.913)
Hansen *J* statistic	0.393	0.114	0.595	0.553	0.377	0.596
Degrees of freedom	1	1	2	1	1	1
Chi-Sqr test (*p* -value)	**(0.530)**	**(0.735)**	**(0.742)**	**(0.457)**	**(0.539)**	**(0.440)**
	First-stage results					
Constant	0.244 (0.000)	0.314 (0.000)	0.248 (0.000)			
Canada dummy	-0.114 (0.000)	-0.121 (0.000)	-0.112 (0.000)			
Log population density 1900	0.004 (0.128)		0.005 (0.292)			
Sex ratio 1900		-0.048 (0.133)	-0.007 (0.863)			
Fraction slaves in 1850	-0.091 (0.001)	-0.101 (0.001)	-0.093 (0.004)			
Partial R-squared of excluded instruments	0.143	0.151	0.166			
F -statistic for joint significance of excl. instr.	6.41	6.2	4.54			
F- test (*p* -value)	(0.003)	(0.004)	(0.007)			
No. of obserbations	59	52	52			

Notes: Two-stage least squares estimations with heteroskedasticity-robust standard errors. Dependent variable is log per-capita GSP or GPP. The instrumented variable is fraction with university degree. The *p* -values in parentheses refer to a *t* -test in the first-stage regression and a *z* -test in the second-stage regression.

Table 5: Two-stage least squares regressions: smaller set of geography instruments

Dependent variable is log per-capita GSP or GPP						
Instrumented variable is:	Log population density 1900		Log population density 2001		Sex ratio 1900	
	Second-stage results					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	10.306 (0.000)	10.292 (0.000)	10.313 (0.000)	10.187 (0.000)	10.730 (0.000)	11.273 (0.000)
Coefficient for instrumented variable	**0.024** **(0.275)**	**0.040** **(0.064)**	**0.015** **(0.557)**	**0.055** **(0.068)**	**-0.316** **(0.186)**	**-0.759** **(0.027)**
Canada dummy	-0.171 (0.036)	-0.174 (0.054)	-0.184 (0.083)	-0.088 (0.460)	-0.248 (0.000)	-0.282 (0.007)
Fraction slaves in 1850		-0.342 (0.092)		-0.514 (0.019)		-0.482 (0.011)
Hansen *J* statistic	2.477	1.248	3.597	1.156	9.027	1.407
Degrees of freedom	2	1	2	1	2	1
Chi-Sqr test (*p*-value)	**(0.290)**	**(0.264)**	**(0.166)**	**(0.282)**	**(0.011)**	**(0.235)**
	First-stage results					
Constant	-16.09 (0.062)		-10.49 (0.060)		0.803 (0.000)	
Canada dummy	-0.84 (0.362)		-2.54 (0.000)		0.052 (0.776)	
Average temperature	0.64 (0.036)		0.47 (0.016)		0.019 (0.752)	
Average temperature Squared	-0.005 (0.044)		-0.004 (0.029)		-0.0002 (0.673)	
Atlantic dummy	1.863 (0.000)		1.444 (0.000)		-0.176 (0.000)	
Partial R-squared of excluded instruments	0.390		0.461		0.256	
F-statistic for joint significance of excl. instr.	13.07		12.32		7.96	
F- test (*p*-value)	(0.000)		(0.000)		(0.000)	
No.of observations	59		60		52	

Notes: Two-stage least squares estimations with heteroskedasticity-robust standard errors. Dependent variable is log per-capita GSP or GPP. Sex ratio is the number of men per woman. The p-values in parentheses refer to a t-test in the first-stage regression and a z-test in the second-stage regression. In columns (1), (3), and (5) only the indicated variables (log population density in 1900 and 2001, and the sex ratio in 1900) are instrumented, one at a time. In columns (2), (4), and (6) also the fraction slaves is instrumented, although the first-stage regression for fraction slaves is not reported.

Table 6: Two-stage least squares regressions: larger set of geography instruments

Dependent variable is log per-capita GSP or GPP

Instrumented variable is:	Log population density 1900		Log population density 2001		Urbanization rate		Fraction with university degree	
	Second-stage results							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	10.305	10.302	10.175	10.189	10.216	10.140	9.451	9.412
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Coefficient for instrumented variable	**0.039**	**0.040**	**0.058**	**0.055**	**0.261**	**0.369**	**3.709**	**3.873**
	(0.054)	**(0.022)**	**(0.056)**	**(0.032)**	**(0.256)**	**(0.095)**	**(0.006)**	**(0.004)**
Fraction slaves in 1850	-0.450	-0.458	-0.556	-0.535	-0.258	-0.275	0.019	0.002
	(0.011)	(0.004)	(0.001)	(0.000)	(0.106)	(0.062)	(0.930)	(0.992)
Canada dummy	-0.186	-0.184	-0.079	-0.090	-0.252	-0.250	0.202	0.220
	(0.027)	(0.030)	(0.518)	(0.428)	(0.000)	(0.000)	(0.244)	(0.212)
Hansen *J* statistic	4.192	4.839	2.046	2.066	4.787	6.989	0.562	4.592
Degrees of freedom	3	4	3	4	3	4	3	4
Chi-Sqr test (*p* -value)	**(0.242)**	**(0.304)**	**(0.563)**	**(0.724)**	**(0.188)**	**(0.136)**	**(0.905)**	**(0.331)**
	First-stage results							
Constant	-12.33	-9.32	-10.20	-7.71	-0.547	-0.378	0.227	0.238
	(0.224)	(0.279)	(0.112)	(0.139)	(0.308)	(0.474)	(0.255)	(0.216)
Canada dummy	-1.26	-1.34	-2.54	-2.60	0.13	0.13	-0.126	-0.126
	(0.194)	(0.110)	(0.000)	(0.000)	(0.081)	(0.073)	(0.000)	(0.000)
Average temperature	0.49	0.37	0.45	0.35	0.04	0.03	0.001	0.001
	(0.199)	(0.258)	(0.056)	(0.066)	(0.031)	(0.074)	(0.804)	(0.861)
Average temperature Squared	-0.004	-0.003	-0.004	-0.002	0.000	-0.0002	0.000	0.000
	(0.198)	(0.281)	(0.081)	(0.116)	(0.069)	(0.167)	(0.714)	(0.779)
Average precipitation	0.046	0.048	0.003	0.004	-0.007	-0.007	-0.0009	-0.0009
	(0.027)	(0.016)	(0.869)	(0.798)	(0.001)	(0.000)	(0.144)	(0.148)
Average number of rainy days	-0.005	-0.008	0.001	-0.0014	0.001	0.001	0.0001	0.0001
	(0.679)	(0.426)	(0.851)	(0.817)	(0.065)	(0.111)	(0.421)	(0.475)
Atlantic dummy	1.33	1.65	1.39	1.65	0.08	0.10	0.030	0.031
	(0.002)	(0.000)	(0.000)	(0.000)	(0.102)	(0.047)	(0.022)	(0.016)
Great Lakes dummy		1.84		1.52		0.10		0.006
		(0.000)		(0.000)		(0.002)		(0.706)
Partial R-squared of excluded instruments	0.455	0.582	0.464	0.606	0.239	0.295	0.14	0.143
F -statistic for joint significance of excl. instr.	8.83	13.49	8.57	14.25	5.24	6.16	2.59	2.55
F- test (*p* -value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.035)	(0.030)
No. of observations	59	59	60	60	60	60	60	60

Notes: Two-stage least squares estimations with heteroskedasticity-robust standard errors. Dependent variable is log per-capita GSP or GPP. The *p* -values in parentheses refer to a *t* -test in the first-stage regression and a *z* -test in the second-stage regression. Both the fraction slaves and the indicated demography variables are instrumented; the first-stage regression for fraction slaves is not reported.