

REGIONAL ECONOMICS: UNDERSTANDING THE THIRD GREAT TRANSITION

Paul Krugman

Revised, October 2019

Memo to young economists: the transition from fiery upstart to ~~old fuddy-duddy~~ elder statesman sneaks up on you. One day, it seems, you're trying to turn everything upside down; the next thing you know you've turned into one of those old guys whose response to any new idea is "It's trivial, it's wrong, and I said it in 1962."

Sure enough, I recently gave the luncheon keynote speech at a Boston Fed conference on America's [growing regional disparities](#), basically providing a break and maybe some inspiration in between presentations by real researchers. The brief talk didn't require a paper, and I am *not* someone who reads prepared speeches. But I thought I might take the occasion to write down a few thoughts on the subject.

Basically, I want to make three points:

1. The regional divergence we've seen since around 1980 probably isn't trivial or transient. Instead, it reflects a shift in the underlying logic of regional growth — the kind of shift that theories of economic geography predict will happen now and then, when the balance between forces of agglomeration and those of dispersion crosses a tipping point.
2. This isn't the first time this kind of transition has happened. In fact, it's the third such shift in the history of the U.S. economy, which went through earlier eras of both regional divergence and regional convergence.
3. There are pretty good although not ironclad arguments for "place-based" policies to limit regional divergence. It's important to realize, however, that the U.S. system already provides huge de facto subsidies to lagging regions. The fact that we're diverging anyway suggests that the economic forces at work are quite powerful.

Petals, fugals, and tipping points

It was thirty years ago today — OK, roughly that — when I realized that some of the modeling tricks underlying the New Trade Theory (or, as younger economists sometimes call it, the old new trade theory) could be applied to the analysis of economic geography. In particular, simple models of monopolistic competition could be combined with simplified representations of transportation costs and factor mobility to produce a "core-periphery" model, which showed both how highly unequal distributions of economic activity across space could arise, and how they could unravel.

For the key insight of the new geography was that there is always a tension between centripetal forces that bring stuff together and centrifugal forces that push it apart. Figure 1, taken from my 1991 paper “Increasing returns and economic geography,” illustrates the point. It envisions an economy consisting of two regions, with some mobile factors and some immobile factors. The horizontal axis shows the share of the mobile factor in region 1; the vertical axis the real wage of that factor in region 1 relative to its wage in region 2:

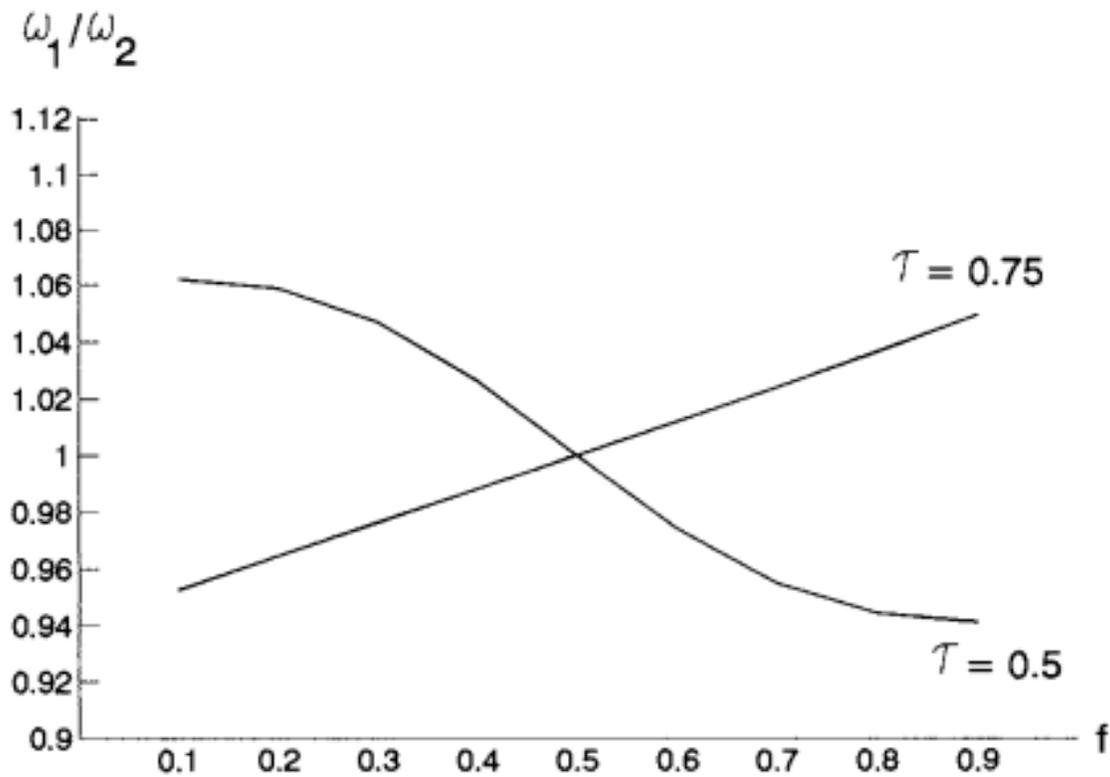


FIG. 1

Depending on the parameters of the economy, the derived schedule could be either downward or upward-sloping. If it was downward-sloping, whichever region started with a higher concentration of mobile factors would offer those factors a lower real return, so concentrations of economic activity would tend to dissipate over time. But if the schedule was upward-sloping, concentrations of economic activity would tend to snowball, with rising regional inequality.

In that paper and subsequent work I followed a “no dormitive properties” rule, trying to derive the forces of agglomeration and dispersion from concrete forces like market access and factor immobility. In real life, however, disembodied externalities like knowledge spillovers and traffic congestion also surely play important roles.

Regardless of the details, the tension between these two kinds of forces suggests that every once in a while quantitative change — a shift in the relative strength of centripetal and centrifugal forces — will bring about a qualitative change in regional dynamics, leading either to a cumulative process of geographical concentration or to an ongoing dissolution of existing concentrations of activity.

In my early work on the subject, I focused a lot on one historical example of such qualitative change: the clear shift toward agglomeration that took place in the middle of the 19th century, as railroads, factory production, and the declining relative importance of agriculture led to the rise of the manufacturing belt and highly localized industries within that belt.

Even at the time, however, this work had something of a steampunk quality; it seemed to be more about the past than the future. Many of the highly localized industries of the past — like the detachable-collars-and-cuffs industry of Troy, New York — are no more. And in general both the dominance of the manufacturing belt and the localization of individual industries seem to have peaked in the early 20th century. By the time the new economic geographers got into the subject, the regional divergence we loved to dwell on was starting to seem like old news, at least in America.

The point is that regional dynamics seem to have reached another tipping point circa 1920 or a bit earlier. Motor vehicles, which reduced the importance of railroad hubs, open-plan factory designs made possible by electrification, which encouraged the movement of manufacturing out of urban centers, and possibly other factors undermined the advantages of concentration, and activity began to disperse around the country.

What we didn't realize — what I, at least, didn't realize — was that even as the new economic geographers were writing about regional divergence as a story about the past, it was turning into a story about the future.

The return of regional divergence

Regional dynamics in the United States didn't change gradually — nor should we have expected a gradual change given our models, which suggest that there are tipping points when quantitative change abruptly changes the whole logic of convergence versus divergence. Sure enough, the long process of narrowing income gaps among U.S. regions, which dated back at least to the early 20th century, came to an abrupt halt circa 1980.

Figure 2 shows a crude indicator of the shift in dynamics (and it happens to be the figure that alerted me to just how profoundly the story had changed.) It shows the ratio of per capita income in Mississippi, our poorest state, to income in Massachusetts, which isn't as rich as New Jersey or Connecticut, but has been near the top all the way back to the beginning of U.S. industrialization. From the 1920s up to 1980, MS was steadily converging on wealthy states, transitioning from a sort of third world nation inside a rich country to a "normal" state. But

then it went into reverse: after rising from 30 to 70 percent of MA income, the state has since slipped back to 55 percent

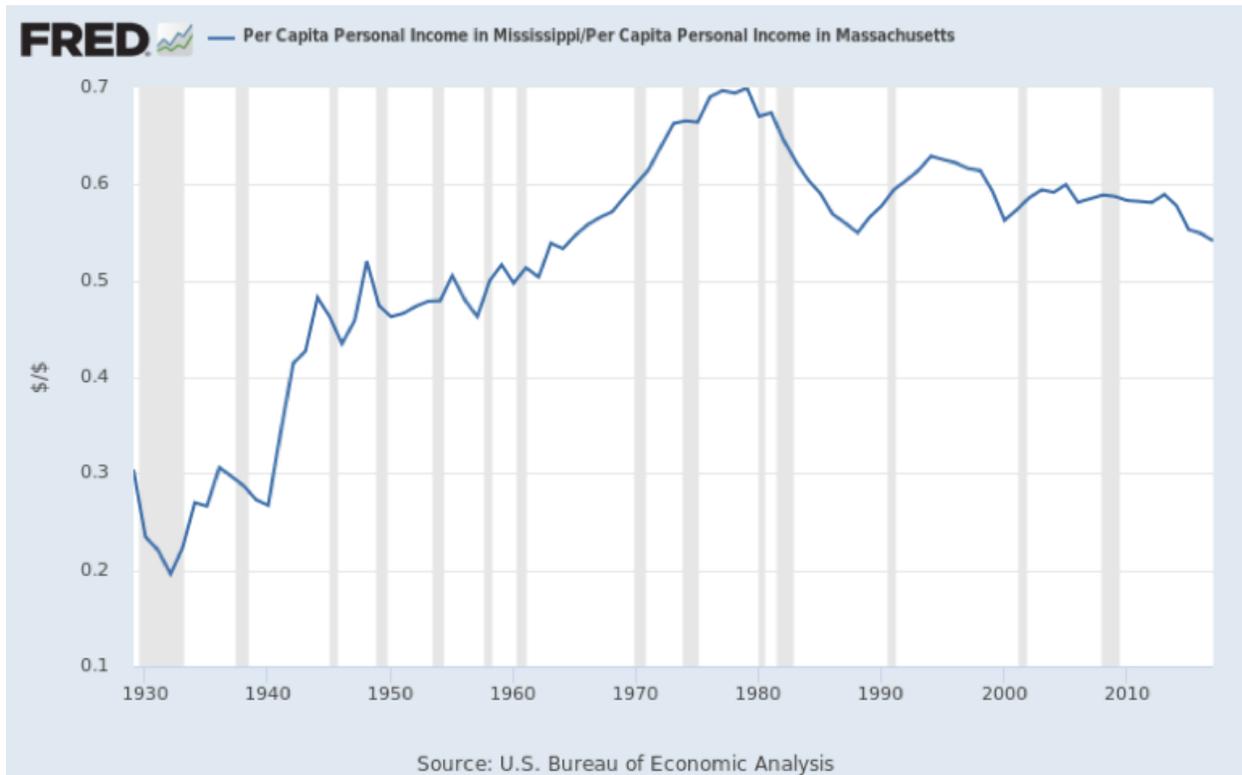


Figure 2

It's convenient to work with state-level data, but much of the return to regional divergence is taking place within states. In particular, big states like California and New York have seen a dramatic parting of the ways between their big coastal cities and the small cities and towns of their interiors. Figure 3 shows the comparison with which Enrico Moretti begins his book on the new economic geography of jobs: Fresno versus San Jose, the heart of Silicon Valley. In the 70s Fresno had 80 percent of San Jose's income, and given low housing costs and other advantages of small-city life, was arguably a better place for many people to settle. But now it's lagging ever further behind, and holds little attraction for the highly educated workers whose movements, as we'll see shortly, seem to be crucial for the new regional divergence.

Of course, focusing on particular pairwise comparisons, while it helps give the story some concreteness and color, runs the risk of cherry-picking. So Figures 4 and 5, both of which plot a state's per capita income in a starting year with growth over an extended subsequent period, provide an across-the-board view of convergence and its absence.

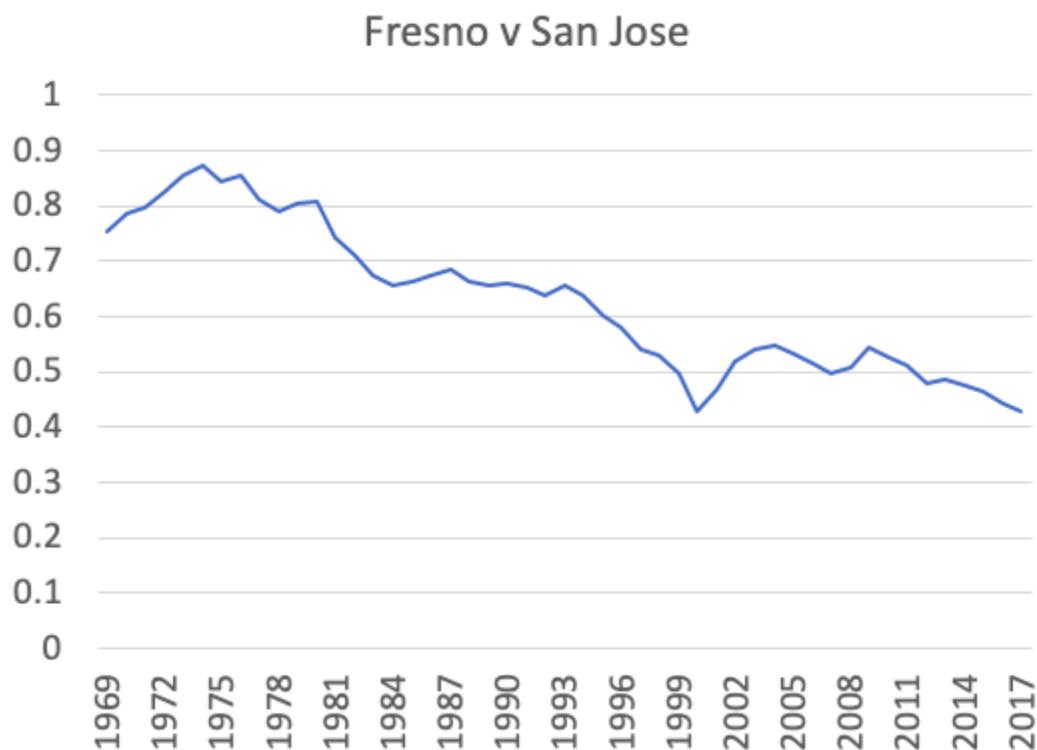


Figure 3

Figure 4 shows what happened over the half-century from 1929 to 1979. There’s certainly no ambiguity about the reality of convergence over that period! In fact, the fit between initial income and subsequent growth is so good as to be almost eerie. Incidentally, the slope of the fitted line — $-.014$ — suggests a “half-life” of regional income differences of just about 50 years.

Figure 5 shows the same comparison since 1979; it obviously looks very different. To be fair, it doesn’t show a clear pattern of divergence; there is at most a slight positive correlation between initial income and the growth rate. I would argue, however, that in a truly neutral environment we *should* expect a downward-sloping relationship, if only because of regression toward the mean, as transitory factors that temporarily raise or lower incomes (like shale oil) go away. The state-level data may also be missing divergence within states. And as we’ll see shortly, other indicators, especially educational attainment, do suggest a systematic process of divergence.

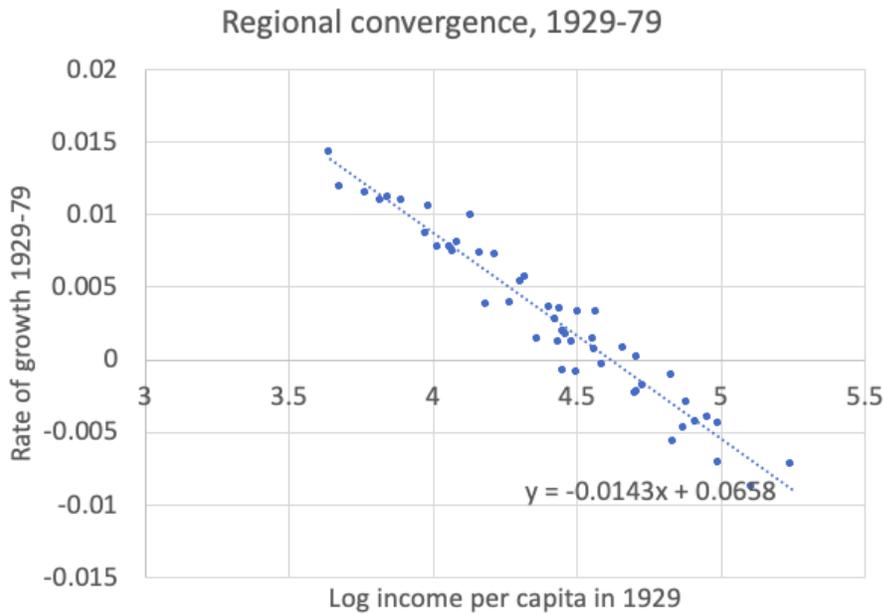


Figure 4

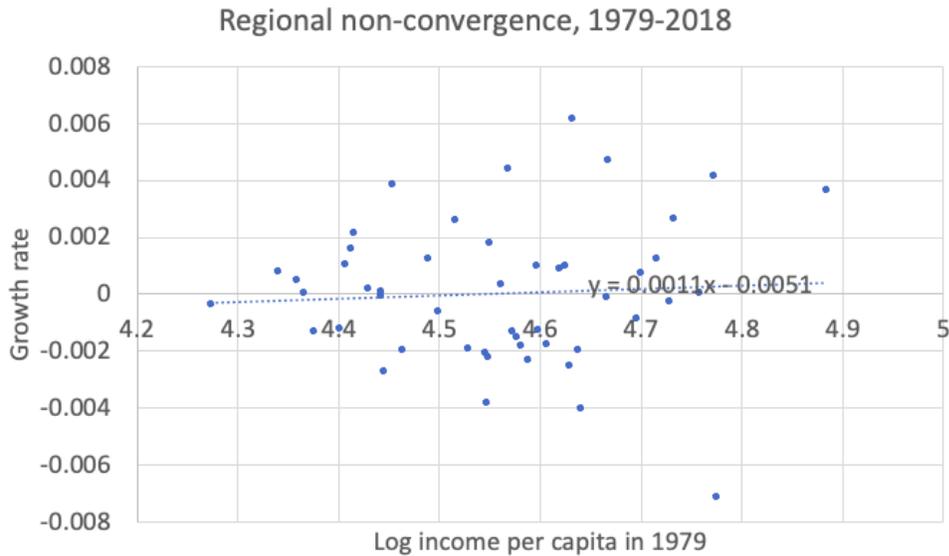


Figure 5

I've already indicated that the apparent shift from convergence to divergence circa 1980 isn't the first reversal we've seen in regional dynamics; in fact, it's the third. Now, we don't have modern-quality data from before 1929. There have, however, been serious efforts to estimate state and regional data back into the 19th century. Figure 6 shows results from one such effort, by Baier et al (2004), who estimated output per worker by region — not quite the same as

income, but closely related — going back to 1840 and continuing to 2000. I show the ratio of output in the East South Central region to output in New England — more or less comparable to the MI/MA comparison in Figure 2.

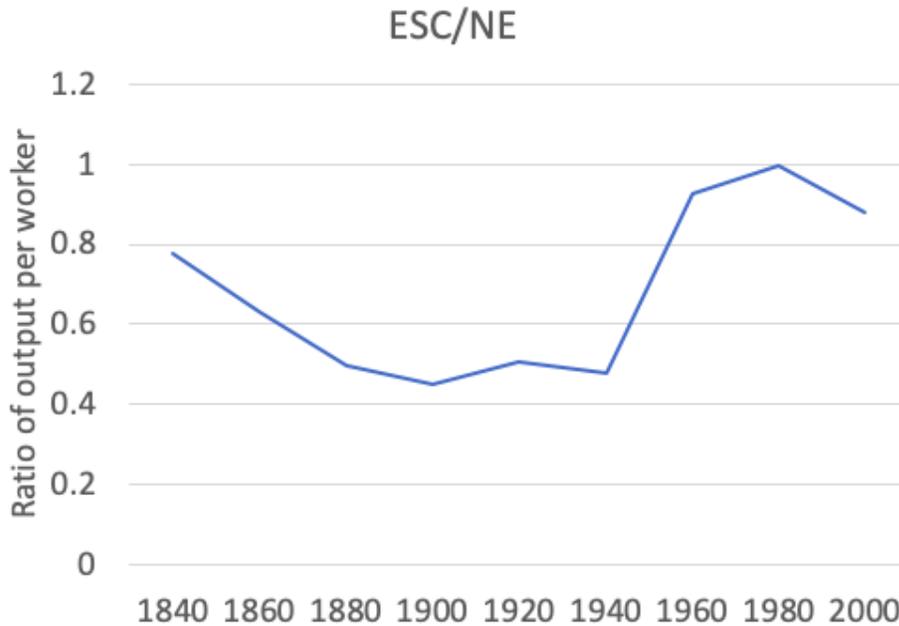


Figure 6

What we see here is divergence more or less from the beginning of large-scale industrialization up to 1900, then the familiar process of convergence from the 1920s to around 1980, and a suggestion of renewed divergence thereafter. As I suggested, we can probably explain the 19th-century and early 20th-century tipping points largely by changing technology, such as the rise first of railroads, then of road transportation.

What caused the third, most recent tipping point, toward regional divergence? The standard explanation, as offered for example by Moretti (2012), is the rise of the knowledge economy: industries that intensively use highly educated workers seem to benefit from geographic clustering of these workers.

And if we look at education levels by state and metro area, we see a clear process of divergence, in which locations that already had high concentrations of college-educated workers have increased their advantage over regions that had fewer such workers to begin with. Figure 7 shows that states with relatively educated populations in 1990 reinforced their lead over the years that followed.

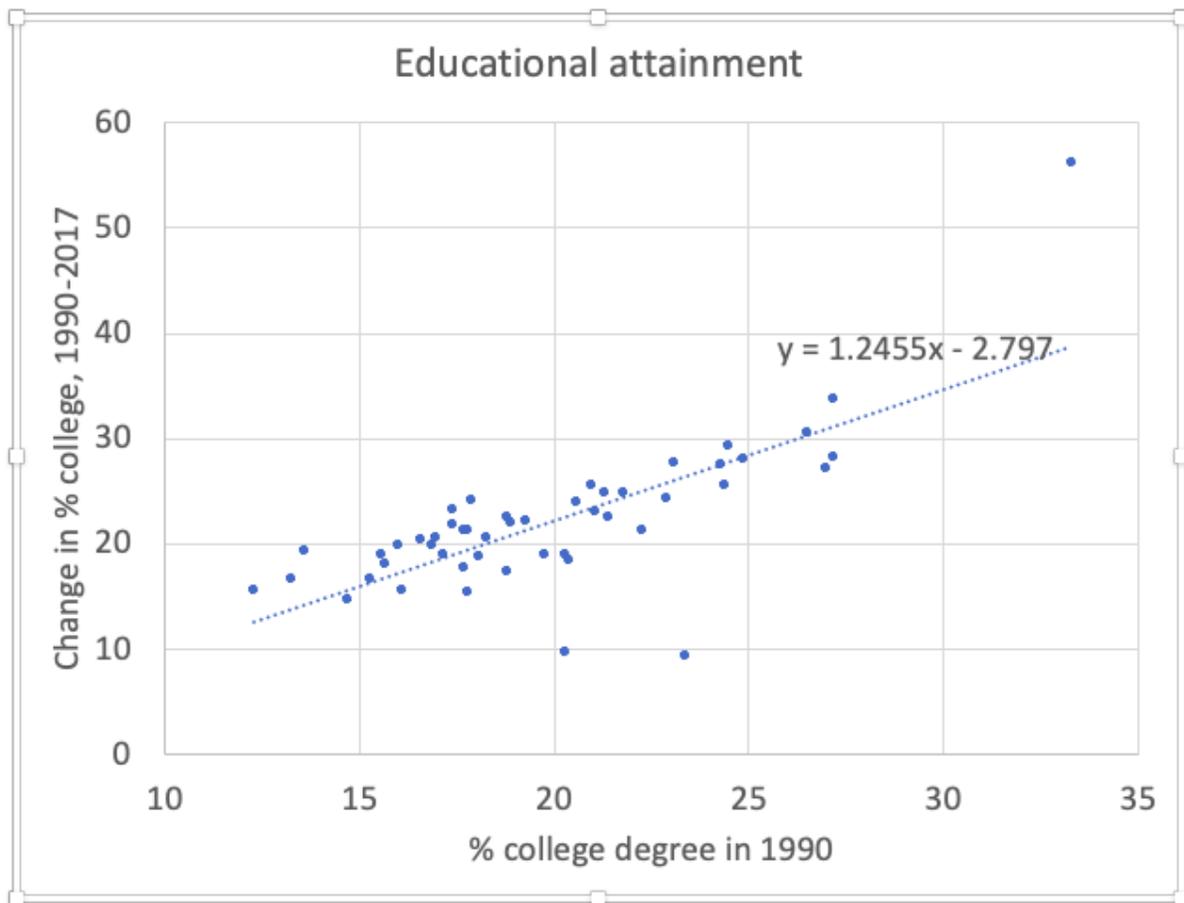


Figure 7

However, I'm also playing with a somewhat different though not contradictory hypothesis: divergence may have been caused in part by reduced communication costs, which make it possible to perform some but not all work remotely. That is, "offsiteing" — not offshoring, which is moving stuff abroad, but just moving it to lower-cost locations in the US — may actually be contributing to divergence.

This may seem paradoxical; doesn't reducing the importance of distance encourage activity to spread out? But here's a sketch of a model: imagine that production requires two kinds of workers, Nerds and Clerks. Nerds benefit from the spillovers that come from being near each other, be it Silicon Valley or Wall Street. Clerks don't. But in the past they had to be located together.

This limited concentration of firms in big metropolitan centers, because firms choosing to locate there had to pay both Nerds and Clerks higher wages to compensate for higher land prices and hence cost of living.

Now develop technology that lets Clerks work remotely. This lets firms move their jobs to lower-cost locations. But it also frees them to move Nerds to high-cost locations, because they can get the spillovers without having to increase Clerk wages.

Assuming Nerds have more education and higher wages, the result is what we see: growing regional disparities in both education and incomes.

Whatever the explanation, we seem to have passed another tipping point, leading to regional divergence. What should or can be done about it?

Divergent America

Regional divergence is an invisible-hand phenomenon, caused by market forces rather than any deliberate policy. And economists normally take as their default position that the market knows what it's doing, that proposals for policies that reject market outcomes must clear a fairly high bar.

In the case of regional dynamics, however, there is no presumption that the market gets it right. Why? Because externalities of some kind are integral to the story. Agglomerations agglomerate because of positive externalities, both information spillovers and pecuniary externalities related to market size, which have welfare consequences in the presence of increasing returns and imperfect competition.

You might think that this creates a presumption that more agglomeration is good — that we should actually encourage regional divergence. But the positive externalities gained by growing regions are gained at the expense of shrinking regions; and there are also negative externalities such as traffic congestion and local air pollution associated with movement of resources to already dense areas. If we sum up all these effects, what's the net impact on the economy? Nobody knows.

Still, economic efficiency isn't the only important issue here. There are also important concerns about social and political cohesion — and on this front regional divergence is clearly causing major problems. As [Austin, Glaeser, and Summers](#) document, America's declining regions are marked by a rising share of prime-age men not working, and rising indicators of social collapse. I find it astonishing to realize that as recently as 1990 [life expectancy](#) in Kentucky and West Virginia was about the same as it was in New York; these days it is 5 years lower.

This economic and social divergence is clearly contributing to political polarization. Figure 8 shows a comparison, produced by the Brookings [Metropolitan Policy Program](#), of the characteristics of Congressional districts in 2008 and 2018 — both years in which Democrats achieved modest-sized majorities.

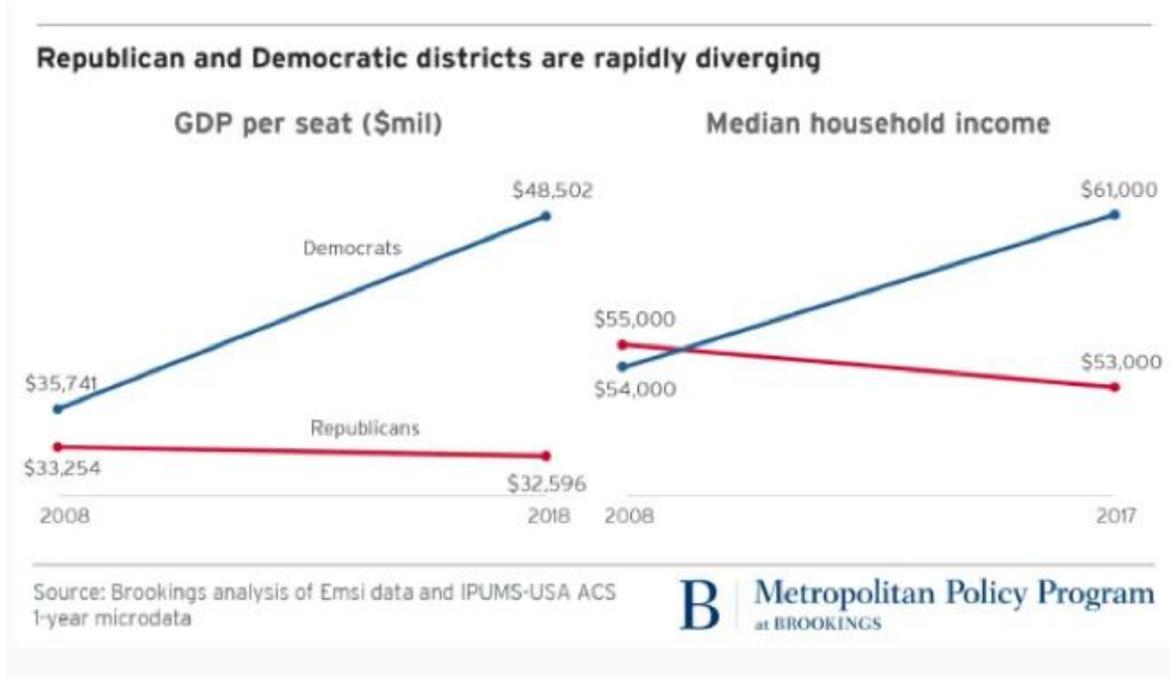


Figure 8

There are obviously other forces at work, but our toxic political environment is clearly not helped by the fact that one party effectively represents the winners from regional divergence, while the other represents the losers.

All of this seems to suggest a role for “place-based” policies to help lagging regions. Other advanced nations have such policies, and the European Union’s “cohesion funds” try to achieve the same thing at a supranational level.

What I don’t think many people in this debate fully appreciate, however, is the extent to which the U.S., while it lacks explicit policies to fight regional inequality, in practice offers huge aid to lagging regions.

I’ve used data from the Rockefeller Institute, which tracks states’ balance of payments with the federal government, to create a summary picture of the U.S. transfer union, shown in Figure 9, which shows per capita federal spending in and receipts from states as functions of the state’s per capita income:

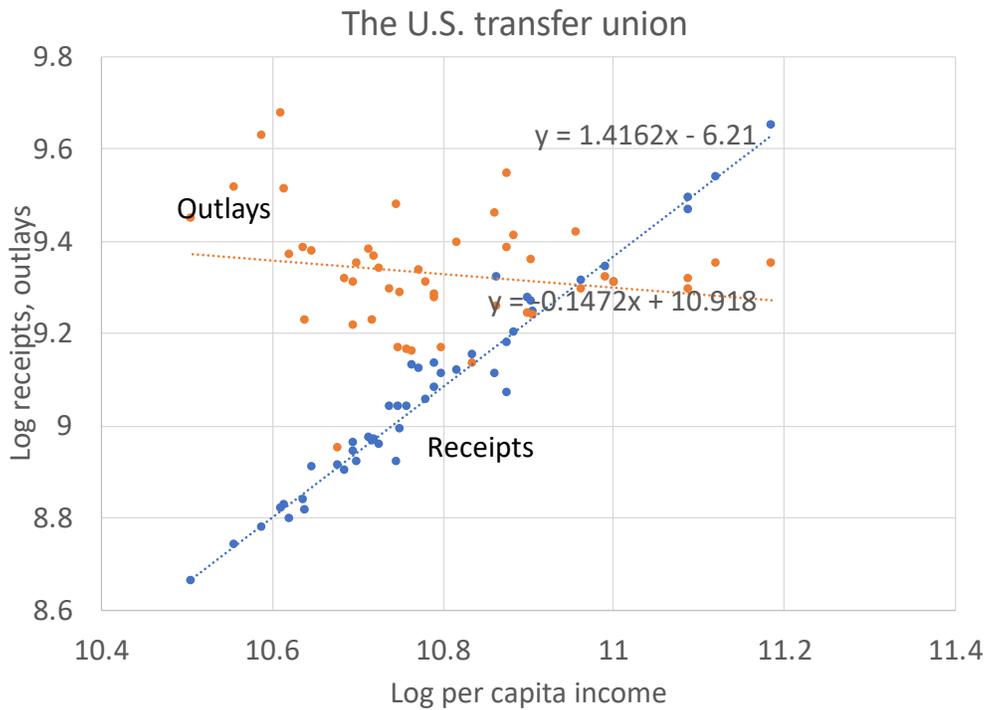


Figure 9

It turns out that federal outlays are slightly declining in state income, reflecting the fact that some though not all federal programs — notably Medicaid and food stamps — are means-tested. Of course this means that federal outlays are much more steeply declining as a share of state income.

Federal receipts, on the other hand, are strongly increasing with state income, with a elasticity well above 1, so that they are a much higher share of rich states' income than of poor states' income.

The combination of these two factors mean that poor states are de facto recipients of substantial, indeed startling, levels of federal aid. Figure 10 shows this de facto aid as a percentage of GDP for the five poorest states; it's on a scale beyond the wildest dreams of EU cohesion policy.

You might be tempted to say that while these large net transfers to lagging regions may support incomes, they don't support employment. But that's not actually true. Boosting the purchasing power of a region's residents supports local service employment, which is actually the majority of employment everywhere. And much federal spending goes on health care, which is a direct source of employment, indeed one of the main remaining sources of middle class jobs

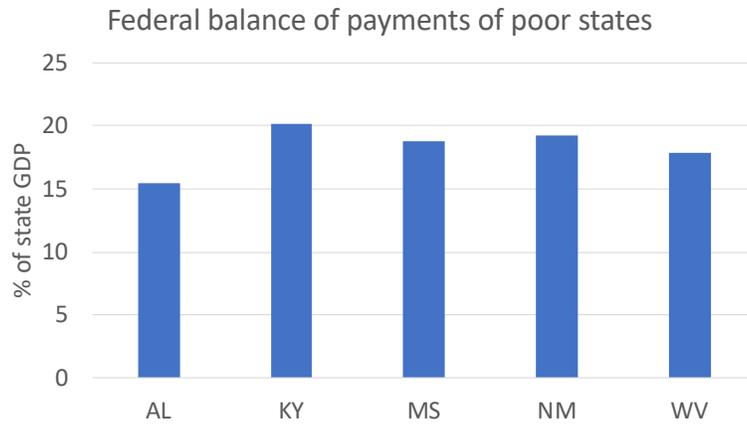


Figure 10

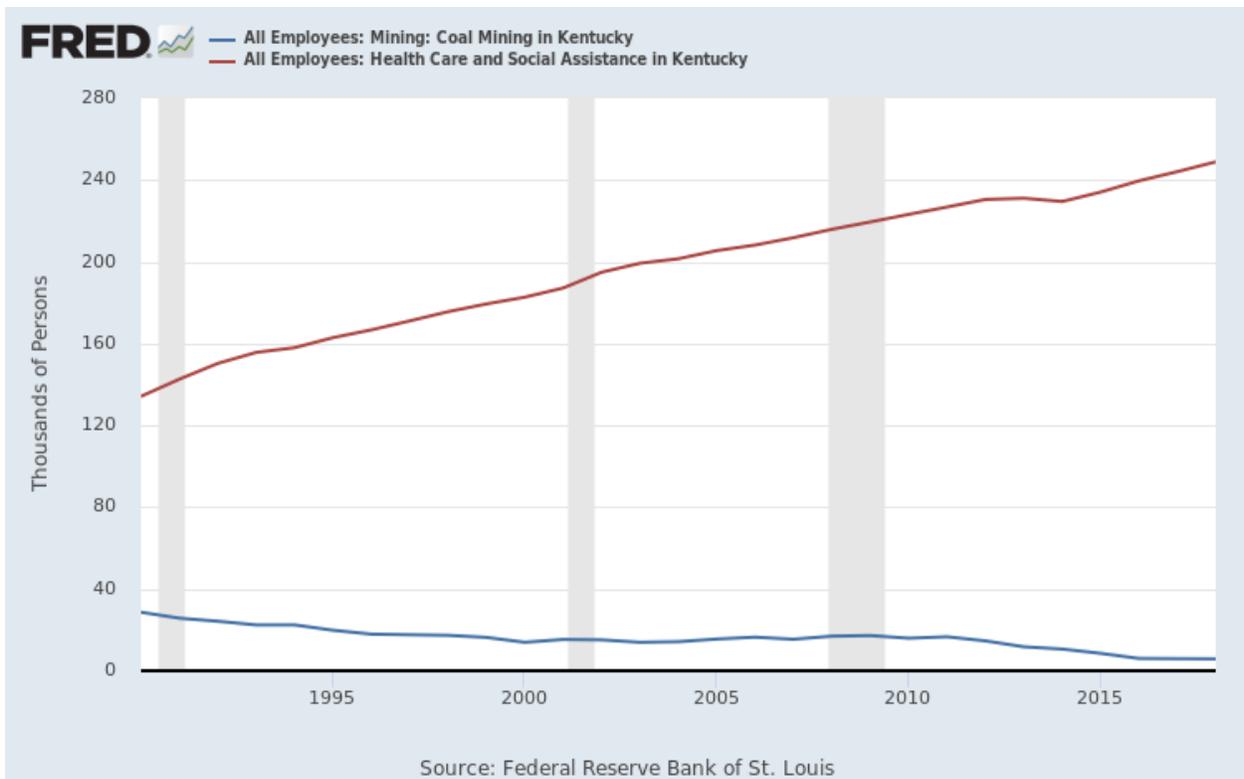


Figure 11

Consider a state like Kentucky, which receives a larger inflow of federal funds relative to GDP than any other state. Kentuckians may still think of themselves as coming from a coal-mining culture, but as Figure 11 shows, coal hardly employs anyone anymore, and the economy's real driver is employment in health care and social assistance — much of it made possible by federal funds.

The point is that the U.S. transfer union is or ought to be a cushion against unequalizing regional dynamics. You might think of it as a sort of spatial equivalent of the automatic stabilizers that play such an important role in smoothing out the business cycle.

And yet regional divergence is happening anyway. This suggests that the economic forces behind this divergence are strong. And I'm sorry to say this, but it inspires considerable pessimism about the ability of policy to reverse the dynamics of divergence. The growing gap between prospering and lagging regions, the social downsides of this gap, and the strong correlation between regional divergence and political polarization, are disturbing. But it's very hard to see any way to stop the process.