No. 2767

EUROPEAN REGIONAL POLICIES IN LIGHT OF RECENT LOCATION THEORIES

Diego Puga

INTERNATIONAL TRADE
EUROPEAN REGIONAL POLICIES IN LIGHT OF RECENT LOCATION THEORIES

Diego Puga, University of Toronto and CEPR

Discussion Paper No. 2767
April 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre’s research programme in International Trade. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre’s publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a Paper should take account of its provisional character.

Copyright: Diego Puga
ABSTRACT

European Regional Policies in Light of Recent Location Theories*

Despite large regional policy expenditures, regional inequalities in Europe have not narrowed substantially over the last two decades, and by some measures have even widened. Income differences across States have fallen, but inequalities between regions within each State have risen. European States have developed increasingly different production structures and European regions have also become increasingly polarised in terms of their unemployment rates. This Paper describes these trends, and discusses how recent location theories can help us to explain them and reconsider the role of regional policies, and especially of transport infrastructure improvements, in such an environment.

JEL Classification: H54, R40, R58
Keywords: inequalities, location, regional policy, transport infrastructure, Europe

Diego Puga
Department of Economics
University of Toronto
150 St George Street
Toronto
Ontario, M5S 3G7
CANADA
Tel: (1 416) 978 4375
Fax: (1 416) 978 6713
Email: d.puga@utoronto.ca

For further Discussion Papers by this author see:
* Prepared for the conference ‘Globalization and the Location of Economic Activities’, Sitges (Barcelona), October 2000. Thanks to Gilles Duranton and Henry Overman for useful comments on early drafts of this Paper and to Ginés de Rus and Jacques Thisse for helpful discussion. Financial support from the Institut Català de Finances and the hospitality of the Instituto de Análisis Económico CSIC are gratefully acknowledged. This Paper is produced as part of a CEPR Research Network on The Economic Geography of Europe: Measurement, Testing and Policy Simulations, funded by the European Commission under the Research Training Network Programme (Contract No. HPRN-CT-2000–00069).

Submitted 26 February 2001
NON-TECHNICAL SUMMARY

Despite large regional policy expenditures, regional inequalities in Europe have not narrowed substantially over the last two decades, and by some measures have even widened. Income differences across states have fallen, but inequalities between regions within each state have risen. European States have developed increasingly different production structures, and European regions have also become increasingly polarized in terms of their unemployment rates.

This Paper argues that recent theories of location can help us explain these trends, and also reconsider the role of regional policies in such an environment. Firms producing in locations with relatively many firms face stronger competition in the local product and factor markets. This tends to make activities dispersed in space. The combination of increasing returns to scale and trade costs, however, encourages firms to locate close to large markets, which in turn are those with relatively many firms. This creates pecuniary externalities, which favour the agglomeration of economic activities.

Reductions in trade or transport costs, by affecting the balance between dispersion and agglomeration forces, can decisively affect the spatial location of economic activities. For high trade costs, the need to supply markets locally encourages firms to locate in different regions. For intermediate values of trade costs, the incentives for self-sufficiency weaken. Pecuniary externalities then take over, and firms and workers cluster together. The price of local factors and the availability of goods, however, tend to increase wherever agglomeration takes place. If this is the case and there is enough mobility, as trade costs continue to fall, rising factor prices simply give an additional kick to agglomeration by inducing immigration. On the other hand, if there is little mobility, for very low trade costs it may be that firms relocate in response to wage differentials. The combination of minimal interregional migration with institutional constraints on interregional (but not international) wage differentials can, however, lead to a rise in income inequalities between regions within each country at the same time as inequalities between countries fall. Further, if agglomeration is not reflected in wage differences, it may show in unemployment rate differences. Since clusters of activity may extend across borders, this can result in clusters of high and low unemployment extending across regions and even across countries. Firms tend to collocate with firms to which they are more closely related and to avoid congestion by moving away from firms to which they are not. This also promotes increasing specialization in a way which is not just driven by traditional comparative advantage considerations.

Whether there is too much or too little agglomeration in the absence of regional policy interventions is not clear. The fact that firms and workers move without taking into account the possible losses for those left behind implies
there may be too much agglomeration. On the other hand, since firms do not fully take into account the benefits they bring for other firms and their impact on aggregate growth and workers when they move, there may be too little agglomeration. Thus, there is no general indication of the direction in which governments should push with regional policies when seeking efficiency. Even in terms of equity, the direction of policy is not obvious. Policies that increase agglomeration may nevertheless make those that remain in poorer regions better off by increasing production efficiency and the rate of growth.

Despite these ambiguities, European regional policies have the explicit aim of reducing regional inequalities. One of the main instruments for this is the improvement of transport infrastructure. It is not obvious, however, that lower transport costs facilitate convergence. Roads and rail tracks can be used to travel both ways. A better connection between two regions with different development levels not only gives firms in a less developed region better access to the inputs and markets of more developed regions, it also makes it easier for firms in richer regions to supply poorer regions at a distance, and can thus harm the industrialization prospects of less developed areas.

New economic geography models not only point out this potential ambiguity in the impact of lower transport costs on less developed regions, they also tell us that the overall effect depends on certain aspects of the economic environment (such as mobility and wage rigidities) and on characteristics of the projects. On this respect, the Trans-European Transport Network will give much of the European Union better access to the main activity centres. The gap in relative accessibility between core and peripheral areas, however, is likely to increase as a result of the new infrastructure, which reinforces the position of core regions as transport hubs. The emphasis on high-speed rail links is also likely to favour the main nodes of the network, and is unlikely to promote the development of new activity centres in minor nodes or in locations in between nodes.
1. Introduction

Profound regional income disparities exist in the European Union. Nearly one quarter of its citizens live in regions eligible to receive assistance under ‘objective 1’ of the Structural Funds, the main instrument of EU regional policy. The criterion for eligibility is to have a Gross Domestic Product per capita below 75% of the EU average. If a similar measure was used for the United States, only two States, Mississippi and West Virginia, containing between them only 2% of the US population, would qualify (Puga, 1999).

There are also large disparities across the regions of the EU in terms of their unemployment rates. In 1996 the 10 regions with highest unemployment rates had twice the EU average unemployment rate. In comparison, differences in unemployment rates across US States are minimal. And differences across European regions are not just the reflection of differences in unemployment rates across countries. Belgium, Finland, France, Germany, Greece, Italy, Spain, and the United Kingdom all have a difference of at least 700 basis points between the unemployment rate of their highest unemployment region and that of their lowest unemployment region.

The growing concern about reducing these regional disparities has made the financial instruments of EU regional policy the fastest-growing component of the EU Budget. The Structural Funds have been allocated €195,000 million (at 1999 prices) for the period 2000–2006. They now account for over 30% of total EU spending (twice the proportion they represented in 1988), and for about 0.4% of total EU Gross National Product. The Cohesion Fund, introduced as a financial instrument in 1993, is to provide another €18,000 million for structural spending in 2000–2006.

Despite this sizable intervention, regional inequalities in Europe have not narrowed substantially, and by some measures have even widened. Over the past fifteen years income differences across Member States have fallen, but inequalities between regions within each Member State have risen. Over the same period, the production structures of EU Member States have become increasingly different. European regions have also become more unequal in terms of their unemployment rates: there are now fewer regions with intermediate unemployment rates than a decade ago, and more regions with either high or low rates.

Recent theories of location can help us explain these trends. Traditionally, international and regional economics have explained income disparities on the basis of differences between regions in their endowments of natural resources, factors of production, infrastructure, or technology. In this context, the removal of obstacles to the movement of goods and/or factors would by itself cause convergence of factor returns and living standards. Yet both casual observation and empirical work in the area show there are relevant forces missing from the traditional analysis, which can widen regional disparities — even without large differences in underlying characteristics — and prevent convergence. The main argument arises from the observation that firms produce more efficiently and workers enjoy higher welfare by being close to large markets, and that large markets are in turn those where more firms and workers locate. This creates a cumulative causation process that tends to open regional differences. Mechanisms of this kind have been described by development

---

1 Defined at level 2 of the Nomenclature of Territorial Units for Statistics (NUTS), a hierarchical classification with three regional levels established by Eurostat to provide comparable regional breakdowns of the Member States of the EU. In 1996 the EU had 77 NUTS 1 regions, 206 NUTS 2 regions, and 1,031 NUTS 3 regions (Eurostat, 1995).
economists and geographers for some time. The main contribution of what has been called ‘new economic geography’ is to bring together, in a common analytical framework, both convergence and divergence forces. The advantage of modelling such forces in a common framework is that we can relate their relative strength to microeconomic conditions, and explicitly study the trade-off between the economic advantages of the clustering of activity and the inequalities that it may bring. Recent location theories can thus help us understand the evolution of regional inequalities during a process of economic integration, and think about the role of regional policy in such an environment. Such an analysis is the object of this paper.

The remainder of the paper is structured as follows. The next section describes the recent evolution of regional inequalities in the EU. Section 3 reviews location theories developed in what has been called the ‘new economic geography’, and uses them to explain the trends described in the previous section. Section 4 looks at European regional policies in light of those theories, and Section 5 focuses on one of the main instruments of such policies, transport infrastructure. Section 6 provides some concluding remarks.

2. Regional inequalities in the European Union

Economic activity is less geographically concentrated in Europe than in the US. Nearly one-half of EU industrial employment is concentrated in a small number (27) of NUTS 1 regions, which account for 17% of the Union’s total surface and 45% of its population. In the US nearly one half of the country’s industrial employment is also concentrated in a small number (14) of States, but these account for much smaller shares of total surface (13%) and population (21%).

Not only is overall activity less concentrated in Europe than in the US, but so are individual sectors. Midelfart-Knarvik, Overman, Redding, and Venables (2000) calculate for EU Member States and for US States an index of spatial separation for individual industries, and divide it by the same index for overall manufacturing. They show that, even after accounting for the wider dispersion of overall industry in Europe, most individual sectors are also more geographically dispersed than in the US.

Midelfart-Knarvik et al. (2000) also compare the industrial structures of EU Member States, and show that these have become increasingly different over the last two decades. In fact, the Netherlands is the only country whose industrial structure has become more similar to the aggregate EU structure. The industrial structures of the four largest EU economies (France, the UK, Italy, and Germany) are relatively similar. However, Germany and Italy are becoming increasingly different from each other as well as from France and the UK. The four Cohesion States (Greece, Ireland, Portugal, and Spain) have also become increasingly different from each other. Spain is now more similar to France, Great Britain, Italy, and Germany than to the other three Cohesion States. Ireland is more similar to Belgium, Denmark, and the Netherlands. Finland and Sweden have remained similar to each other but have become increasingly different from the rest.

---

2 This spatial separation index is calculated as the production-weighted sum of all bilateral distances between locations, which takes value zero if all production takes place in one place and increases as production becomes more geographically dispersed.
The increasing sectoral specialisation of EU Member States has followed an uneven pace. Figure 1 (from Midelfart-Knarvik et al., 2000) plots an average index of national specialisation relative to the EU for four groups of Member States (the initial Member States, the 1973 entrants, the 1980s entrants, and the 1995 entrants). This shows little change in specialisation during the 1970s, and then a phase of rising specialisation which has been particularly acute since the early 1980s for the 1980s entrants and since the mid-1990s for the 1995 entrants.

Given that production is less geographically concentrated in Europe than in the US, one might expect a more even distribution of income per capita as well. In fact, the opposite is true: differences in income across European regions are much wider than across US States. In 1992 the ten best-off regions had a GDP per person equal to 1.6 times the Union’s average and 3.5 times that of the ten worst-off regions (at NUTS 1 level). By comparison, the ten best-off US States had a GDP per person equal to 1.2 times the US average and 1.5 times that of the ten worst-off States.

European regions experienced a clear convergence in income per capita up until the late 1970s, when convergence came to a sudden stop (see, amongst others, López-Bazo, Vayá, Mora, and Suriñach, 1999, Canova and Marcel, 1995, Marcel, 1994, Neven and Gouyette, 1995, and Rodríguez-Pose, 1999). Figure 2 (from de la Fuente and Vives, 1995) shows the clear exhaustion of regional income convergence after the 1970s.

Additional information can be gathered by decomposing income inequalities, as measured by a Theil index, into inequalities across countries and inequalities across regions within each country. Figure 3 (from Esteban, 1999) shows that during the first half of the 1980s income inequalities across EU Member States accounted for about one half of overall regional inequalities, and inequalities
Figure 2. Dispersion of regional income in Europe

Figure 3. Decomposition of a Theil index of income inequalities across European regions
across regions within the same State for about another half. Since then income inequalities across States have fallen by 25%, but regional inequalities within States have risen by 10%. As a result, nowadays most regional income inequalities in Europe are within rather than across Member States.

Looking at inequality indices provides a useful, but partial, picture of the evolution of regional inequalities across European regions. For instance, using summary statistics alone one cannot discriminate between a situation in which regions roughly maintain their relative position over time, and another one in which the shape of the distribution changes little but the relative positions of particular regions change significantly over time. One way to get at this is by looking at geographical maps illustrating the whole regional distribution for different points in time.

Figure 4 graphs the GDP per capita relative to the EU average of the NUTS 2 regions that made up the EU in 1987 for that year and for 1995. The map shows only small changes in the relative income level of individual regions. This is in contrast with the notable changes that appear when same map is drawn for relative unemployment rates in 1986 and in 1996, as shown in Figure 5.

More details can be learned by constructing transition probability matrices that track changes over time in the relative position of regions within the distribution. This is an exercise that a number of authors, including Quah (1999), have undertaken. The transition probability matrix at the top of Table 1 is an update on that work, which reports transitions between the 1987 and 1995 distributions of GDP per capita relative to the European average (only regions that were part of the EU at the beginning of the period are included in the calculation). The table also gives two additional pieces of information. The first column gives $n$, the number of regions that begin their transitions in a given state. The second column gives the classes that divide up the state space. The main diagonal of the matrix shows the fraction of regions that were in the same range of the distribution in 1995 as in 1987. The large numbers on this diagonal show the strong persistence of relative regional income levels. Reading along the top row of the matrix, we see that of the 30 regions that in 1987 had a GDP per capita below 0.6 times the EU average, 83% remained in the same range in 1995, while 17% saw their relative income rise to between 0.6 and 0.75 times the EU average, and none moved higher up in the distribution than this. The proportion of regions that experienced little relative change is very high for all ranges of the distribution, although regions with a 1987 GDP per capita of between 0.6 and 0.75 times the EU average (second row from the top) showed more mobility: 21% saw their income fall below 0.6 times the average, and 32% saw it rise to between 0.75 times the average and the average.

Again it is insightful to compare the distributions of GDP per capita and unemployment rates. Up until the mid-1980s differences in unemployment rates across European regions were very stable, with changes in regional labour forces just offsetting ongoing changes in regional employment (see chapter 6 in Layard, Nickell, and Jackman, 1991). The transition probability matrix at the bottom of Table 1 (from Overman and Puga, 1999) shows that this stability no longer holds. It reports transitions between the 1986 and 1996 distributions of unemployment rates relative to the European average. Reading along the top row of the matrix, we observe a strong persistence for regions starting with an unemployment rate below 0.6 times the European average: by 1996, 81%...
Figure 4. Regional GDP per capita in Europe
Figure 5. Regional unemployment rates in Europe

Source: Overman and Puga (1999)
remained below 0.6 times the European average, 19% had an unemployment rate between 0.6 and 0.75 times the average, and none had a relative unemployment rate higher than that. The next row down tells us that of those regions with an initial unemployment rate of between 0.6 and 0.75 times the European average, 26% remained in that range, while 52% saw their unemployment rate fall below 0.6 times the average. Jumping to the bottom row we also see strong persistence amongst the regions with highest unemployment rates. However, regions with unemployment rates between 0.75 and 1.3 times the European average (third and fourth rows from the top) experienced much greater mobility: regions with initial unemployment rates between 0.75 times the average and the average ended up almost equally distributed across the four intervals between 0 and 1.3 times the average.

The contrast between changes in relative GDP per capita and changes in relative unemployment rates can be seen most clearly by comparing the numbers on the diagonal in the third, fourth, and fifth rows of both transition matrices. This shows that, while European regions exhibited a quite strong persistence in their relative income levels, they have experienced a polarisation of regional unemployment rates since the mid 1980s. As a result of regions with intermediate unemployment rates moving towards both extremes of the distribution, there are now more regions with either very high or very low unemployment and fewer regions in between. The extent of this polarisation of unemployment rates can be measured using the generalisation by Esteban, Gradín,
and Ray (1999) of the polarisation measure of Esteban and Ray (1994). Between 1986 and 1996, the polarisation of the unemployment rate distribution into a group of high and a group of low unemployment regions increased by 37%, from 0.096 to 0.131 (Overman and Puga, 1999).

Looking back at Figure 5 we can see that the polarisation of unemployment rates has a strong geographical component. Comparing the maps for 1986 and 1996 we can see clearly defined clusters of high and low unemployment emerging, and these extend not just across regions but often also across countries. As a result national borders have become increasingly blurred. Overman and Puga (1999) study this process more formally and show that indeed, even after controlling for regions’ characteristics, there is a strong spatial effect. Unemployment rates are more similar across neighbouring regions than across regions in the same State or with similar characteristics. Regardless of their initial position, regions also experienced similar changes to those of neighbours. And surprisingly, this similarity not only persists after controlling for common characteristics, but is also equally strong across neighbours in the same country as across neighbours in different countries. While Figure 4 also shows a strong neighbourhood component in relative GDP per capita (see Quah, 1996, for a formal analysis), the spatial pattern there is better described by a centre-periphery gradient (see Keeble, Olford, and Walker, 1988) rather than by localised transnational clusters.

How can we make the polarisation of unemployment rates consistent with the relative invariance of GDP per capita? GDP per capita and unemployment rates are related to each other. In fact, one can decompose GDP per capita as follows:

\[
\frac{\text{GDP}}{\text{population}} = \frac{\text{GDP}}{\text{employment}} \times \frac{\text{employment}}{\text{labour force}} \times \frac{\text{labour force}}{\text{working-age population}} \times \frac{\text{working-age population}}{\text{population}}
\]

\[
= \text{productivity} \times (1 - \text{unemployment rate}) \times \text{participation rate} \times \text{working-age pop. share}.
\]

One possible explanation for the polarisation of unemployment rates not being reflected in substantial changes in GDP per capita is that this polarisation could be driven mostly by changes in the size of the labour force in each region, with small changes in employment. Then changes in unemployment rates would not have been reflected in income per capita due to changes of opposite sign in the ratio of the labour force to population. Overman and Puga (1999) show that this is not the case. Most regions that ended up with relatively low unemployment in 1996 had relatively high employment growth over the previous decade. Similarly, regions that ended up with relatively high unemployment generally had below average employment growth. It is employment changes that have driven high unemployment regions to their high rates and low unemployment regions to their low rates. Labour force changes, on the other hand, have worked against polarisation, but have been insufficient to prevent this.

This means that unemployment changes must have been closely matched by productivity changes in the same direction. That is, regions that have lost employment have seen productivity rise and regions that have gained employment have seen productivity fall. As a result, there have been only small changes in the distribution of GDP per capita.

---

4Esteban (1999) decomposes inequalities in GDP per capita, as measured by a Theil index, into the components in the above identity. He finds that, while differences in productivity are still the largest component, differences in unemployment rates are increasingly important.
This still leaves open the question of what is driving productivity changes of the same sign as unemployment changes. It is probably not differences in sectoral specialisation, given that inter-regional productivity differences are mostly common across sectors (Esteban, 2000). Another, rather mechanical, possibility is that changes in unemployment are moving in and out of employment workers with below-average productivity. However, something else may be driving productivity changes. This is an issue we still know little about. What we do know is that, over time, there has been convergence in productivity, divergence in unemployment rates, and little change in GDP per capita across European regions.

In order to help interpret the evolution of regional inequalities across Europe described in this section and to better understand how it might be affected by further integration and regional policies, we now turn to a brief review of some recent location theories.

3. Forces driving the new economic geography

Economic theory has traditionally explained differences in production structures mainly through differences in underlying characteristics (endowments of natural resources, factors of production, infrastructure, or technology), which make space itself uneven. In this framework, economic integration leads regions to specialise according to their comparative advantage. Comparative advantage, while relevant, provides a weak explanation for the remarkable spatial concentration of activity. In order to explain the uneven geographical distribution of economic activities across areas with similar endowments, we must consider increasing returns to scale (Scotchmer and Thisse, 1992, call this the ‘folk theorem of spatial economics’). Models of trade with increasing returns and imperfect competition provide an explanation as to why regions without significant comparative advantage with respect to each other can develop different production structures on the basis of their different market access.

The implications of these models for location, and the effects that reductions in trade or transport costs have on it, are formalised by Krugman and Venables (1990). They start by assuming that there are two regions: a large ‘core’ region and a small ‘peripheral’ region. There are two factors of production, which are mobile across sectors but immobile across regions. The core region has larger factor endowments than the peripheral region, although both have the same relative endowments — hence there is no comparative advantage in the traditional sense. There are two production sectors. One of these sectors is perfectly competitive and produces a freely tradeable homogenous commodity under constant returns to scale. The other sector is imperfectly competitive and has firms producing differentiated manufactures under increasing returns to scale; They label this ‘manufacturing’.

It is hardly surprising that at equilibrium the core has a larger manufacturing sector than the periphery. The interesting finding is that, for finite positive trade costs, the core’s share of industry is larger than its share of endowments. It is therefore a net exporter of manufactures. This effect is
known as the ‘market access’ or ‘home market’ effect. Furthermore, each region’s share of industry changes non-monotonically with trade costs. This is best seen by considering a process of gradual reduction in trade costs between the two regions from prohibitively high to none, as depicted in Figure 6. The vertical axis is the share of industry in each region, and the horizontal axis plots trade costs (zero represents free trade, one represents trade costs equal to the producer price of the product). In this example region 1 (the core) is assumed to have 60% of total endowments of the two factors.

With high trade costs, firms sell mainly — but not only — in their local market. Then if a region had many more firms relative to its market size than the other region, the greater competition or crowdedness of that market would lead some local firms to exit, reducing differences in the size of industry. As a result, each region’s share of industry is close to its share of endowments.

Economic integration increases the share of sales that each firm makes in the other region, thereby weakening the effect of more local competitors on each firm’s market share. Yet increasing returns imply that the larger sales of firms producing in the core give them higher profits. If more firms enter in response to those profits, the size of industry in the core rises above its share of world endowments.

However, as the size of industry in the core increases so does demand for local factors. For low

\[ \text{Figure 6. Trade costs and location in Krugman and Venables (1990)} \]

---

\[ ^{5}\text{Davis (1998) introduces transport costs for the good produced under constant returns in a model similar to the one described here, and shows that this results in no equilibrium trade in this good, and hence no market access effect. Fujita, Krugman, and Venables (1999) show that re-establishing equilibrium trade in the constant returns good by having this being differentiated across countries re-establishes the market access effect. Davis and Weinstein (1999) find evidence of market access effects for Japanese prefectures. For OECD countries they do not when they focus on pure national market size, but they do when they consider a richer measure of market access (Davis and Weinstein, 1998).} \]
trade costs, rising factor prices start driving some firms out of the core, so further integration starts reducing its share of industry. As the two regions approach costless trade it is increasingly factor price differences that determine location, so differences in both nominal and real wages between them tend to disappear, while each region’s share of industry tends to go back to its share of overall endowments.

This type of trade model with imperfect competition highlights the fundamental ambiguity of the effects of economic integration or reductions in transport costs on the relative attractiveness of core and peripheral regions. However, like traditional theory, it still explains differences in production structures through differences in underlying characteristics — in this case in terms of exogenously given market size. The main contribution of the so-called ‘new economic geography’ (or ‘new geographical economics’) is the formalisation of mechanisms by which, even a priori very similar regions, can end up with very different production structures and income levels.⁶ The simplest such mechanism arises when one introduces some mobile source of demand in the framework just described.⁷

---

**Endogenous core-periphery structures with labour mobility**

Krugman (1991) shows that the interaction of labour migration across regions with increasing returns and trade costs creates a tendency for firms and workers to cluster together as regions integrate. If some factors are mobile between regions, then the pressure put on those factors by the concentration of economic activities will be eased. Factor mobility can make the supply of factors sufficiently elastic that small differences in the size of industry across regions can build up. Even if regions are a priori identical, they can become endogenously differentiated into an industrialised core and a deindustrialised periphery.

Krugman (1991) considers two regions and two sectors: one monopolistically competitive, the other perfectly competitive. There is no mobility of workers across sectors, and only industrial workers are mobile across regions.⁵ Finally, the two regions are assumed a priori identical in every respect, including their endowment of immobile factors.

To understand the forces at work it is useful to consider the following thought experiment. Suppose that both regions have equal shares of industry but that, for some reason, one firm decides to move production from one region to the other. How does this affect the profitability of firms in the region receiving the firm? The presence of one more firm puts additional pressure on the product and labour markets, and tends to make the firm go back to where it came from. If there was

---

⁶The review of the new economic geography in this section only aims to highlight some of the main mechanisms at work and not to survey the literature; it draws from a more comprehensive survey by Ottaviano and Puga (1998). Martin (1999c) and Neary (2000) provide critical reviews of this work. See also Fujita and Thisse (2000) for a complementary survey of how this approach relates to work based on either technological externalities or spatial competition.

⁷Alternatively, factor accumulation can play a similar role (Baldwin, 1999).

⁸Introducing intersectoral mobility in Krugman’s 1991 model (but retaining some other immobile factor) does not substantially change the results, but, by making labour supply to this sector more elastic, strengthens the tendency of industry to agglomerate (Puga, 1998).
no migration, this would be the end of the story and regions would remain identical. However, the rise in the share of goods produced locally in the region receiving the firm (hence free from trade costs) and the rise in local labour demand and wages tend to attract more workers. This increases local expenditure and eases pressure in the labour market, so tends to attract more firms. Whether the overall effect of entry is to increase the profitability of local firms (encouraging further entry), or to lower that profitability (leading to exit), depends on parameters of the model, and in particular on how integrated regions are.

Figure 7 plots the share of industry in each region during a gradual reduction in trade costs between the two regions. This shows that when trade costs fall below some critical value, whichever of the two regions gets a slight advantage will build on it. The presence of more industry attracts more workers due to higher wages and a lower cost of living. The presence of more workers in turn attracts more firms through a market access effect. And this self-reinforcing mechanism leads to an endogenous differentiation of the two a priori identical regions into an industrialised core and a deindustrialised periphery. Since regions are a priori identical, which region becomes the industrialised ‘core’ and which the ‘periphery’ when they endogenously differentiate is not determined by endowments, but instead by history — and, in extensions to this model (see, for instance, Ottaviano, 1999), by expectations.

**Endogenous core-periphery structures without labour mobility**

The cumulative causation mechanism modelled by Krugman (1991) relies on the assumption that, when a region does relatively well in attracting firms, it is able to attract more workers on the basis of higher wages and better access to a wider range of goods. Blanchard and Katz (1992) show
that in the US there is such an adjustment process working through regional migration. That is not the case in Europe, where adjustment to changes in regional fortunes takes place mostly through participation decisions (Decressin and Fatás, 1995).

Migration rates in Europe are low in comparison with those of the US (Bentivogli and Pagano, 1999). They are also low by historical standards, specially when compared with the much higher migration rates that characterised Europe in the 1960s. The Single European Act was set to create a single market for goods and workers in the EU, yet only 1.5% of EU citizens live in a Member State different from where they were born. Even within European countries, migration across regions remains small. There is some disagreement as to whether this is the result of people being reluctant to move, the incentives being insufficient, or barriers to migration being too high.

Low mobility in Europe is often blamed on language and cultural barriers, or on Europeans being immobile per se. Yet culture cannot easily explain low migration rates within European countries, which in most cases do not have substantial internal language barriers. Nor can it explain the dramatic fall in migration rates since the 1960s.

Migration rates could also be low because of lack of incentives. This may seem an unlikely explanation given that, as discussed in Section 2, there are large differences across European regions both in terms of income and unemployment rates. Yet in many European countries interregional wage differences for similar jobs are relatively small, an issue we shall return to below. Nevertheless, Bentolila (1997) argues that in the case of Spain the importance of wage convergence for the decline in migration is likely to be small, given the almost nil elasticity of migration with respect to interregional wage differentials (Bentolila and Dolado, 1991). Faini, Galli, Gennari, and Rossi (1997) argue that the same is true for Italy. Nominal wage equalisation there was largely achieved at the end of the 1960s following a union agreement to abolish regional wage differentials. Yet the decline in migration has continued steadily.

Even with similar wages, differences in the probability of finding a job still result in interregional differences in expected income. Empirical studies tend to find zero responsiveness of interregional migration in Europe to unemployment differentials (Bentivogli and Pagano, 1999). But de la Fuente (1999) finds that when one considers more sophisticated measures of the probability of finding a job than the usual one minus the unemployment rate, people are much more responsive to differences in this probability. This suggests that one of the main reasons why there is not more migration in Europe is that the high average unemployment rate makes the probability of finding a job anywhere too low to make it worthwhile to move (Braunerhjelm et al., 2000). Also, inefficiencies in the job-matching process may result in a much lower probability of being hired outside one’s own region. Faini et al. (1997) cite evidence from Casavola and Sestito (1993) showing that Italy’s unemployed rely largely on the networks of family and friends to find a job. Such networks are much less effective outside the region where the unemployed are located. Faini et al. (1997) find this to be a key reason for the low propensity to migrate across Italian regions: those amongst the unemployed in the Mezzogiorno who rely more heavily on family and friends to find a job are also much less likely to migrate. Monfort and Ottaviano (2000) formalise the link between the efficiency

of the job matching process and agglomeration and suggests that, if labour markets function more efficiently in areas where firms agglomerate, then differences in unemployment rates across regions may be very persistent.

A low probability of finding a job outside one’s own region, be it due to high unemployment rates or inefficiencies in job-matching is clearly only part of the story. When the unemployed are asked as part of the Spanish labour force survey whether they would take a job away from their place of residence if offered one, only 33.2% answer yes (in 1997, down from 44.6% in 1987) (Pérez and Serrano, 1998).

Rising incomes may have made people more sensitive over time to amenities in their place of residence (de la Fuente, 1999). Family and government support could also help explain the decline in the propensity to migrate (Attanasio and Padoa-Schioppa, 1991). According to this argument, young people, who typically constitute the bulk of migrants, are less willing to move because rising family income has strengthened family support when unemployed. While this argument makes sense, there are forces working in the opposite direction. Migration is a costly investment and rising family income can help finance it more easily. Faini et al. (1997) find that in Italy this second effect tends to dominate: higher household income (proxied by the family’s employment rate) increases rather than decreases the propensity to migrate. At the same time, distortions introduced by certain government benefits may be important. Antolín and Bover (1997) find that in Spain those amongst the unemployed receiving benefits (as proxied by being registered unemployed) are significantly less likely to migrate.

Finally, the costs of housing transactions and the difficulties of finding a rented accommodation can also play a prominent role in discouraging interregional mobility. Oswald (1998) argues that high house ownership rates in Europe are linked to low mobility and high unemployment rates. In the case of the UK, McCormick (1997) discusses how rising housing prices in the faster growing regions tend to discourage interregional migration.

What do new economic geography models have to say about the relationship between trade costs and regional inequalities if migration does not occur, be it due to lack of mobility or to the incentives to move not materialising?

Venables (1996) addresses this question by arguing that firms benefit from being close to each other not only because of linkages working through the supply of labour and demand for goods from each others workers, but also because of direct input-output linkages amongst themselves. He shows that vertical linkages between upstream and downstream industries, when both of them are imperfectly competitive, can play a role equivalent to that of labour migration in endogenously determining the size of the market at different regions.

Krugman and Venables (1995) pick up this argument, and formalise it in a framework more directly comparable to that in Krugman (1991). Cumulative causation arises in a similar way in both papers. In Krugman (1991) an increase in the number of firms in a location increases demand for the output of local firms through the expenditure of the workers attracted from other regions. In Krugman and Venables (1995) there is no interregional mobility, so workers must be drawn from other sectors instead, and the higher demand comes from expenditure on intermediates by the newly arrived firms. In addition to this demand linkage, in Krugman and Venables (1995) there is
a cost linkage arising from the saving in trade costs on a larger fraction of their intermediate inputs by firms in the larger market (these demand and cost linkages can be seen as a formalisation of Hirschman’s, 1958, backward and forward linkages).\textsuperscript{10}

The relationship between transport costs and agglomeration, however, becomes more complex without labour migration. If agglomeration does not open interregional wage gaps (which in the context of Krugman and Venables, 1995, simply requires that at equilibrium all countries keep some constant returns to scale production), things work as in Krugman (1991): a fall in trade costs below some critical level leads industry to concentrate in a single region. However, if the concentration of industry in some regions tends to make their wages relatively higher, the lack of interregional labour mobility can make the relationship between regional integration and industrial agglomeration non-monotonic, so close enough integration can trigger the industrial take-off of less developed regions.\textsuperscript{11}

This Ω-shaped relationship between transport costs and agglomeration is studied by Puga (1999), with a unified framework in which both interregional migration and input-output linkages may drive agglomeration.\textsuperscript{12} He finds that the lack of interregional mobility introduces two main differences in the relationship between economic integration and regional inequalities, which are illustrated in Figure 8.

First, comparison of outcomes with and without interregional migration shows that agglomer-
ation gets an extra kick from the relocation of workers towards locations with higher real wages. On the other hand, if workers do not move, wage differences persist and act as a dispersion force by increasing production costs for firms producing in locations with relatively many other firms. This dispersion force can moderate agglomeration and sustain non-extreme equilibria in which all regions have industry, even if in different proportions. Thus the lack of interregional mobility both postpones agglomeration in a process of regional integration and weakens it when it happens.

One straightforward implication of this is that lower mobility in Europe together with higher barriers to trade can play an important role in explaining why non-agricultural employment is less geographically concentrated in Europe than in the US but income disparities are wider across EU regions than across US States.

The second difference introduced by the lack of mobility is the non-monotonicity in the relationship between trade costs and agglomeration. With zero trade costs each firm finds no advantage in locating close to the rest of industry and locates in the region with lowest wages; therefore, if wages are increasing in industrial employment, for trade costs sufficiently close to zero agglomeration in one region cannot be an equilibrium. Krugman and Venables (1995) illustrate this with examples in which for low trade costs some firms relocate from the industrial agglomeration to regions with lower wages, but not to the extent of allowing full convergence between a priori identical regions. Puga (1999) shows that, more generally, for high trade costs firms are split between the identically endowed regions to meet final demand; for intermediate trade costs regional disparities open up as some regions attract more industry than others — but not necessarily to the extent of absorbing all of industry; for low trade costs agglomeration unravels as the share of industry in regions with lower wages increases gradually (early entrants look for lower prices of immobile factors relative to more industrialised regions; later, as a critical mass of firms is created in some sectors, more firms move in to exploit forward and backward linkages).

In a multi-country version of Krugman and Venables (1995) motivated by the experience of European integration, Puga and Venables (1997) find that the formation of a customs union has location effects that benefit the union as a whole, as firms find it increasingly advantageous to serve the common market from inside (Baldwin and Venables, 1995, call this a ‘production shifting’ effect). However, such gains are unevenly distributed between the integrating countries, with early stages of integration opening up differences in production structures and real income levels between the integrating countries, and later stages leading to convergence.

Puga and Venables (1996, 1999) study the details of the phase of spreading development triggered by cost differences when there are more than two regions. This shows that development may not be a smooth process by which poorer countries gradually and simultaneously catch up with the rich. Instead, the propensity of economic activity to agglomerate may cause the division of regions into a group of industrialised regions and a group of less industrialised regions. Development as well as economic integration initially encourages further agglomeration into a subset of countries. This accentuates wage differences between industrialised and non-industrialised countries. At some point this wage gap becomes unsustainable, and industry starts to spill over to low wage economies. As this process continues, relocating firms begin to benefit from the forward and backward linkages to other firms, and a ‘critical mass’ is reached by some region. At this
point that region undergoes rapid industrialisation, accompanied by an increase in wages. Further growth causes the process to repeat itself, so industry spills over, in a series of waves, from one region to another.

The $\Omega$-shaped relationship between integration and agglomeration of many of these models could be interpreted to suggest that, due to low interregional labour mobility, European integration may by itself cause regional convergence both in terms of real wages and of production structures. However, the ability of poorer regions to catch up in this context relies on integration going far enough (during intermediate stages of integration the model predicts possibly large interregional real wage disparities), on similar endowments in terms of skill and technologies, and on a flexible response of wages to changes in industrial employment. Making sure these criteria are met is essential to achieve convergence.

With respect to the last point it is worth noting that institutional constraints limit interregional wage differentials within European countries. German reunification led to a sharp reduction in regional wage differences between the Eastern and Western Länder. According to Akerlof, Rose, Yellen, and Hessenius (1991), wages in the East rose by 42% between the first quarter of 1990 and October of that same year. This rise occurred while migration was small and falling, and was largely due to strong union bargaining for wage equalisation on the grounds of preventing large scale migration, and the perception on the part of workers that higher unemployment in the East did not justify lower wages.\textsuperscript{13} In Italy and Spain, wage setting at the national level limits the responsiveness of wages to regional economic conditions (on this respect, see Jimeno and Bentolila, 1998, who stress the impact on the regional structure of wages in Spain of wage floors set at the national sectoral level). The rebates on social security contributions introduced in Southern Italy to create some labour cost differences have done so along a very coarse geographical partition (North-South) and without improving the responsiveness to local conditions, so have not offset rigidities in the wage setting process. They are now being eliminated following an agreement with the European Commission.

Puga (1999) suggests that the combination of minimal interregional migration with wage setting at the national sectoral level may help understand the rise in income inequalities between European regions within each country over the last 15 years at the same time as inequalities between countries have fallen. In the models discussed above, when agglomeration does not get reflected in wage differentials, agglomeration increases monotonically with integration. If the structure of wages in Europe reflects differences in local conditions between countries more than differences between regions within each country, further European integration could reinforce the current trend: peripheral countries catching up in their average income to core countries, while poorer regions keep falling behind (see also Faini, 1999, for a formalisation of this argument which explicitly incorporates union behaviour). If agglomeration does not get reflected in wage

\textsuperscript{13}The equalisation of social security contributions also contributed towards rapid wage equalisation. While the one to one conversion rate between Ostmarks and Deutschmark has also been blamed, this is not a compelling reason. As Bean (1992) puts it, ‘w]ould things really have been different if a conversion rate of, say, 10 Ostmarks to the Deutschmark had been chosen? While much of the East German industry might have been viable at that rate, it would have remained so only as long as East German workers were willing to accept the corresponding low real wages. [...] At the end of the day the problem is that East German workers want a West German standard of living, which East German capital is not presently capable of delivering.’
differences, it may get reflected instead into differences in unemployment rates. Since clusters of activity may extend across several administrative units, this can result in clusters of high and low unemployment extending across regional and even national borders.

One stylised fact we have not related to the models discussed so far is diverging industrial structures. However, in order to study the relationship between trade costs and sectoral agglomeration, one needs to consider a more detailed input-output structure.

**Agglomeration and regional specialisation**

Krugman and Venables (1996) consider a setup like that in Krugman and Venables (1995), with one main difference: the two production sectors are imperfectly competitive, and firms in each sector sell and buy a higher proportion of intermediates to and from firms in the same sector than to and from firms in the other sector. This introduces an important difference with respect to Krugman and Venables (1995): if one more firm locates in a region, the beneficial cost and demand linkages affect more intensely firms in the same sector, while the increased product and labour market competition harms firms in both sectors equally (Henderson, 1974, uses a similar argument to explain city specialisation). As a result, a fall in trade costs below some critical level leads each region to become specialised in the production of one sector.

Venables (1999) extends the model in Krugman and Venables (1996) to a continuum of imperfectly competitive sectors and a perfectly competitive sector. He then asks what proportion of sectors will be located in each of the two regions when agglomeration occurs. With just two sectors the answer was one industry in each region, this meaning that both regions have the same income levels. But with many industries the division need not be half and half. One region can have more industries than the other, this leading to real income differences between regions. What Venables shows is that there are bounds to sustainable regional differences, and that the maximum share of total industry that one region can capture first increases and then decreases during a process of regional integration. However, because within those bounds the actual division of sectors between regions is indeterminate, there are strong incentives for each region to try to secure the maximum possible number of sectors.

Most trade models predict that specialisation will increase with reductions in trade or transport costs. What distinguishes new economic geography models in this respect is the fact that even similar regions or countries can develop very different specialisation patterns. This provides a justification for the increasing differences in production structures across different core countries as well as across different peripheral countries in Europe described in Section 2.

Having discussed how some recent location theories can help understand the evolution of regional inequalities in Europe, we now turn to the policies that aim to affect that evolution.

4. European regional policies

The main instrument of EU regional policy, the Structural Funds are articulated around three ‘objectives’. Objective 1 is ‘promoting the development and structural adjustment of regions whose development is lagging behind’. Objective 2 is ‘supporting the economic and social conversion of
areas facing structural difficulties’. Objective 3 is ‘supporting the adaptation and modernisation of policies and systems of education, training and employment’. Of the €195,000 million (at 1999 prices) allocated to the Structural Funds in the 2000–2006 budget, 69.7% is allocated to objective 1. The NUTS 2 regions eligible under objective 1 are those with a Gross Domestic Product per capita below 75% of the EU average. The Cohesion Fund is to provide another €18,000 million over the period 2000–2006, in this case for Greece, Ireland, Portugal, and Spain.

Despite the articulation of European regional funds under different objectives, it is not always clear what the aims are. The first and fundamental issue for a good design of regional policies is to define the objectives clearly. Do we want a homogenisation across space of certain aggregate measures (such as income per capita, unemployment or employment rates, or health and education indicators)? Or is the objective one of personal fairness, similar people having similar opportunities in different places?

The two objectives are related but do not necessarily go in the same direction. The case of the UK is a good example. Since the early 1980s there has been a dramatic divergence of both average real and average nominal earnings across UK regions. However, individual earnings for people with similar characteristics (such as sex, education and experience) have steadily converged. The fundamental reason underlying aggregate divergence in the British case is the combination of three factors: a rise in the premium to education, an increase in the geographical concentration of more educated people in London and the South-East, and a convergence of the premium to education which was initially lower in these two regions. Had the premium to education and interregional differences in this remained stable, the so-called North-South divide would have decreased rather than increased (Duranton and Monastiriotis, 2000).

If, at least in part, the objective of territorial policies is the spatial homogenisation of certain aggregate measures it is important to realise that there are limits to this process that will be more or less important depending on which mechanisms for convergence are considered acceptable. Seeking full spatial homogeneity would obviously be absurd. The location theories reviewed in the previous section highlight the potential efficiency gains from the clustering of activity. However, if there is sufficient mobility, agglomeration need not be incompatible with a convergence of income and unemployment rates. With little mobility, on the other hand, reducing regional inequalities may involve giving up some of the gains from agglomeration (see Martin, 1999a, b, for an elaboration of this argument). The idea that regional policies should aim to disperse activity with respect to the market outcomes indicates a presumption that market outcomes are characterised by too much concentration. However, there are forces pushing in opposite directions. There is a tendency for too much agglomeration, since when firms and workers move they do not take into account the possible losses for those left behind. And there is a tendency for too little agglomeration, since when firms and workers move they only take into account their own benefit and not the benefits they bring for other firms and their impact on aggregate growth. Thus, there is no general indication of the direction in which governments should push with regional policies when seeking efficiency. Even in terms of equity, the direction of policy is not obvious. Martin and Ottaviano (1999) show in the context of a new economic geography model with endogenous growth, that policies that increase agglomeration may nevertheless make those that remain in poorer regions
better off by raising the rate of growth. Combes and Linnemer (2000) show in the context of a location model à la Hotelling that policies that induce asymmetries in the location of production may help consumers everywhere by intensifying competition and lowering prices.

Turning to the instruments of European regional policy, similar proportions of the Structural Funds allocation are devoted to training, subsidies to enterprises, and infrastructure investments. New economic geography models have little to say with respect to training, since most of these models do not incorporate human capital accumulation. One of the few conclusions they offer on this respect is that, insofar training facilitates innovation and knowledge diffusion, it can both increase aggregate growth and reduce regional inequalities (Martin, 1999b). Training can also be important in the context of increasing sectoral specialisation. Most trade models predict that specialisation will increase with integration and reductions in transport costs. What distinguishes new economic geography models on this respect is the observation that the pattern of specialisation may not be driven by traditional comparative advantage considerations but instead by self-reinforcing agglomeration. While specialisation will increase the need for schemes designed to help workers withstand adjustment, the focus of such training schemes is not straightforward. It can be hard to distinguish permanent from temporary shocks on a certain sector. And the skills needed to survive each type of shock are very different. While sector-specific skills can help withstand temporary shocks by building a regions ‘depth of comparative advantage’, it is general skills that will help workers adapt in the face of permanent shocks.

Regarding subsidies to enterprises, new economic geography models analyse in detail some of the externalities that arise from the location of individual firms, which affect those in other regions. As usual in the presence of externalities, this implies that there is a trade-off between centralisation or coordination to internalise these externalities as well as those that might arise from competition between governments, and decentralisation to exploit the better knowledge of local jurisdictions, and gains from competition and diversity. The multiplicity of equilibria in new economic geography models also highlights that the distribution of activities we observe is just one of many possible outcomes, so that individual governments may have incentives to push for certain outcomes that favour them but not necessarily other governments or aggregate welfare. Also, governments may be tempted to contain industrial change. Looking back at Figure 1, the 1970s stand out as a period in which neither countries already in the European Community nor new entrants experienced a clear increase in their industrial specialisation. One possible interpretation is that specialisation was stifled by the particularly strong industrial policy interventions that characterised Europe over that decade. Nevertheless, the role of governments and the games they play in new economic geography models has not yet received the necessary attention. Hence a detailed discussion of aids is beyond the scope of this paper (see Besley and Seabright, 1999, Braunerhjelm et al., 2000, for two recent studies of state aids from a wider perspective).

However, the focus of new economic geography models on the relationship between transport costs, agglomeration, and regional inequalities makes these models particularly appropriate to study the role of improvements in transport infrastructure.
5. Transport infrastructure as a regional policy instrument

The European Commission sees transport infrastructure improvements as playing ‘a key role in efforts to reduce regional and social disparities in the European Union and in the strengthening of its economic and social cohesion’ (Commission of the European Communities, 1999). The Commission is thus promoting the development of a Trans-European Transport Network (TEN-T). This includes 14 priority projects, endorsed by the Essen European Council of December 1994, and a large number of smaller projects. Preliminary estimates put the total cost at over €300,000 million (at 1993 prices). Projects developed as part of the TEN-T are eligible for substantial Community support, particularly in the ‘Cohesion’ Member States (Greece, Ireland, Portugal, and Spain). The EU budget for 1995–1999 devoted a total of €2,300 million to the TEN-T. In the 2001–2006 budget the figure was doubled to €4,600 million. National governments will also devote substantial sums to transport infrastructure. Public infrastructure investments in Spain, mostly for transport, are expected to amount to 2.7% of GDP annually over 2000–2006.

An explicit motivation behind these large investments is the view of transport infrastructure as one more input into the production process. Thus, increasing the stock of infrastructure in less developed regions, like increasing any other stock of capital, is bound to help these regions grow closer to more developed ones. This view underlies a large number of econometric exercises estimating aggregate production functions in which public capital enters as an input. Early exercises of this sort (Aschauer, 1989) reach the implausible result that one more unit of government capital pays for itself by means of higher aggregate output in less than a year. In the US and in other industrialised countries the rate of public investment in infrastructure fell after the late 1960s and the rate of productivity growth fell shortly afterwards. However, concluding from this that the productivity slowdown was caused by a slowdown in infrastructure investment is quite a leap. Even if there is a causal relationship, it is not clear in which direction it runs. Perhaps it was having a sustained period of growth that facilitated large infrastructure investments and not the other way around. Moreover, even if infrastructure investments did increase the rate of growth, this does not imply that further investments will; particularly given that, when building a transport network, the most beneficial links tend to be built first. More recent contributions in the macroeconomics literature, exploiting cross-sectional detail and using careful econometrics find more modest returns to infrastructure investment (see Gramlich, 1994, de la Fuente, 2000, for reviews).

Transport infrastructure is indeed a production input; nearly all goods require some form of transport — although transport accounts for less than 5% of costs in most sectors. But one should not forget that its role in transporting goods and people across space gives transport infrastructure very different properties relative to other forms of capital. The remainder of this section discusses some of those properties.

What are the effects of an improvement in transport infrastructure? The first impact of a transport project comes from the construction expenditure. Given the sums involved, this is not negligible. A transport project also generates costs and revenues associated with its operation. Further, it also has a direct impact on regions affected, typically by reducing the cost and increasing
the quality of transport between them; this in turn induces changes in the total number of journeys undertaken, and in the way in which these are split between different modes of transport. All of these effects, together with the environmental impact, are normally considered as part of the economic evaluation of projects.

It is therefore instructive to look at the cost-benefit analysis of some recent transport infrastructure projects in Europe. A noteworthy aspect is the shift of emphasis in European transport infrastructure investment from roads to high-speed rail. In 1996–1997, both modes of transport represented similar shares of TEN-T investment, each taking about €15,000 million. Almost two thirds of railway investment went to high-speed routes. The balance is now tipping towards rail. So much so that, on the whole, rail investments within the TEN-T framework are expected to amount to €185,000 million (at 1993 prices), more than twice as much as roads, and over six times as much as airports.

This in itself reflects one of the characteristics of high-speed rail: its high sunk infrastructure cost, as compared to conventional rail and roads. Track costs of the only high-speed rail line built in Spain prior to the TEN-T (Madrid-Sevilla) were about €6.5 million (at 1993 prices) per kilometre. De Rus and Inglada (1997) perform an ex-post evaluation of this line. Since the Madrid-Sevilla line started its operations in 1992, it has taken a large fraction of passenger transport away from both car and air transport. Yet de Rus and Inglada’s (1997) careful cost-benefit analysis arrives at a negative net present value of the project of at least €2,300 million (at 1993 prices) for an infrastructure investment of roughly the same amount. More broadly, Nash (1991) casts a skeptical view of the overall benefits versus costs of high-speed rail proposals.

Traditionally cost-benefit analysis does not try to assess the impact of transport projects on regional economic development. This approach of looking only at those activities more directly related to the project can be enough to get an accurate evaluation, provided that two conditions are met. First, that distortions and market failures are not significant, so that private and public valuations are not too different. Second, that the changes in levels of activity induced by the project fade away fairly rapidly as we move away from those activities more closely related to it. However, these conditions are often not met. There has been increasing realisation throughout economics that wide ranges of economic activities may be affected by market failure and distortions. And the type of cumulative causation mechanisms modelled by the new economic geography can make the effects of a project be amplified rather than dampened as they spread through the economy.

For high-speed rail these sorts of additional effects may not be very important. High-speed rail lines are generally not suitable for the transport of goods, and are thus unlikely to have much effect on the location of industry. Even traditional rail accounts for a relatively small fraction of goods transport in the EU: about 8% (in terms of tons × kilometre), as compared with about 40% in the US. At the same time, high-speed rail may have larger effects on the location of business services and headquarters. Duranton and Puga (2001) suggest that the resulting increase in the ability to provide business and headquarter services to remote locations may lead to the concentration of business services and headquarters in a few large centres, and a shift from sectoral to functional urban specialisation. There is indeed some informal evidence that the construction of the Paris-Lyon high-speed rail line led to the relocation of headquarters from Lyon to Paris. In Spain, there are
concerns that the Madrid-Barcelona high-speed rail line may reinforce the process of headquarter relocation towards the capital (Vives, 2000).

Road infrastructure, being more heavily used to ship goods, is likely to have larger effects on the spatial allocation of production, and hence on regional inequalities. These sorts of effect have recently been used to justify the use of transport infrastructure investments as one of the main instruments of regional policy. According to Europe’s Committee of the Regions ‘transport should not be judged on strictly economic criteria (economic viability), but considered in the context of a broader socio-economic and environmental analysis. In this respect, it is important to highlight that if lack of or inadequate basic services are not offset by an efficient transport network that diminishes the adverse effects of such deficiencies by providing access for the population of isolated or disadvantaged regions, this will serve to increase depopulation and reduce economic activity, thus hampering returns on transport investment. The result is a vicious circle in which the growing lack of supply generates a growing lack of demand and vice versa’ (Committee of the Regions, 2000).

One should not forget, however, that roads generally have lanes going both ways, and high-speed train lines also have tracks going both ways. A better connection between two regions with different development levels not only gives firms in a less developed region better access to the inputs and markets of more developed regions. It also makes it easier for firms in richer regions to supply poorer regions at a distance, and can thus harm the industrialisation prospects of less developed areas.

New economic geography models not only point out this potential ambiguity in the impact of lower transport costs on less developed regions, they also tell us that the overall effect depends not just on characteristics of the projects, but also on certain aspects of the economic environment. For instance, if there is little interregional migration, and if wages do not vary much between regions — even when regions differ widely in their attractiveness to firms — then investments in infrastructure can do little to help poorer regions catch up, and may even widen their lag with respect to richer regions. This is not just a theoretical possibility. Faini (1983) has convincingly argued that the reduction in transport costs between Northern and Southern Italy in the 1950s deprived firms in the Mezzogiorno from the protection they had previously enjoyed, and accelerated the deindustrialisation process in Southern Italy.

Of course, the effects of a transport project on the spatial allocation of production depend crucially on the specific details of the project. New economic geography models explicitly capture the role of transport infrastructure in facilitating trade, and in doing so make it possible to discriminate between similarly-sized projects facilitating different types of trade, or interconnecting places in different ways. Martin and Rogers (1995) were the first to focus explicitly on the role of infrastructure in a new economic geography framework (and to take into account the costs of financing it). They distinguish between projects that facilitate trade within a region, and those that facilitate trade between regions. While improvements in interregional infrastructure may harm rather than help peripheral regions — for the reasons discussed above — improvements in local infrastructure lack such negative effects. Another important distinction is between hub-and-spoke interconnections, in which places are connected by routes going via a common centre or hub, and multilateral connections in which places are connected pairwise with routes of similar quality.
Puga and Venables (1997) and Fujita and Mori (1996) study this distinction in the context of new economic geography models. They show that hub-and-spoke networks promote agglomeration in the hub of the network, as firms located there face lower transport costs to spoke locations than firms in one spoke to another. Furthermore, they also tend to trigger disparities between spoke regions.

There is evidence that the road and high-speed rail networks being built as part of the TEN-T programme are likely to generate this kind of hub effect. Gutiérrez and Urbano (1996) study changes in accessibility as a result the trans-European road network. Their results are illustrated in Figure 9, with greater accessibility represented by lighter shades. Much of the area of the EU gains better access to the main activity centres as a result of the trans-European road network, and some of the biggest absolute changes in accessibility take place in peripheral regions which start with very low levels of transport infrastructure. However, the gap in relative accessibility between the areas with the best and the worst initial accessibility increases as a result of the network.

In a related study, Vickerman et al. (1999) analyse how the TEN-T for high-speed rail will change the relative locational advantage of different parts of the European continent. The three maps in Figure 10 are time-space maps from that paper. They are drawn in such a way that the distances between locations are not proportional to their physical distance as in topographical maps, but proportional to the travel times between them. A short travel time between two points results in their presentation close together on the map, while points separated by long travel times appear distant on the map. The base map on the left-hand side assumes a constant travel speed of 60 km./h. in all directions so that it looks like a standard topographical map. The map in the centre is drawn instead for travel times by rail using the existing European rail network in 1993, and the map on the right-hand side using the European rail network planned for 2010 with all the new high-speed rail lines being built as part of the TEN-T. These maps show that, while Europe as a whole ‘shrinks’ in terms of the time it takes to get from one place to another by train, areas

---

They measure accessibility using the GDP-weighted sum of the impedances (quality-adjusted times of travel) to all nodes.

Figure 9. Accessibility by road, 1992 and 2002

Source: Gutiérrez and Urbano (1996)
at the geographical core of the continent get drawn together the closest. This is, once again, the hub effect at work. When major cities get connected through high-speed rail lines, cities located more centrally get better access to nearly everywhere whereas in more peripheral locations the improvement is felt mostly in the access to nearby locations. This has led the European Commission to acknowledge that ‘[i]n transport policy, cohesion countries stand to gain in absolute terms from trans-European networks but not necessarily in relative terms’ (Commission of the European Communities, 1996).

One of the characteristics that distinguishes high-speed rail from other forms of transport is its strong nodal aspect. Too many stops and a high-speed train stops being high-speed. In many cases intermediate stops continue to be served by slower local trains. Another distinguishing characteristic of high-speed rail is the large magnitude of sunk infrastructure costs relative to operating costs. Both these characteristics are important for the spatial organisation of production. It is a well known result in location economics (see Beckmann and Thisse, 1986) that places that lie in between the main nodes of a network are unattractive locations for production. Further, transport technologies that exhibit increasing returns to scale, as is the case with high-speed rail, are unlikely to promote new centres of production even on nodes of the network.15

The two maps in Figure 11 (also from Vickerman et al., 1999) plot daily accessibility surfaces, drawn so that the height of the surface at a particular point of the map is proportional to the population that can be reached from that particular point by a return rail trip made during a working day with some minimum stay in the destination (essentially population within five hours of train travel). This seems a particularly appropriate accessibility indicator for high-speed rail travel, since this mode of transport is specially relevant for the business service sector. The map on the left-hand side is drawn for the existing European rail network in 1993, and the map on the

---

15 The Hakimi Theorem (Hakimi, 1964, Guelicher, 1965) states that any location minimising total transport costs for a producer must be a node of the network and not an intermediate point. In an extension to this theorem, Louveaux, Thisse, and Beguin (1982) show that increasing returns to scale in transport technologies exclude from the cost-minimising set of locations those nodes that are not already large markets for the product in question, or important input sources in the case of resource-driven activities.
right-hand side for the European rail network planned for 2010. It can be seen that the introduction of new high-speed lines leads to strongly non-monotonic changes in accessibility, and creates large differences in accessibility within small distances. While the population that can be reached within five hours by train greatly increases, differences on this respect between the core and the periphery as well as between the main cities and smaller cities or rural areas increase. Only cities that are nodes of the high-speed rail network gain accessibility, while the areas in between nodes and those not on the network or at its edges do not. Further, the growth in accessibility of cities in the European core is several times larger than that of cities at the European periphery.

The last three figures presented illustrate how changes in transport infrastructure lead to changes in accessibility. To find the effects on regional inequalities requires one further connection, linking changes in accessibility with changes in the spatial allocation of economic activities. A few recent papers have built on new economic geography models to study this connection. Venables and Gasiorek (1999) develop a methodology for quantifying the general equilibrium effects of specific projects, which is meant to complement more traditional cost-benefit analysis. They calibrate a new economic geography model with data for EU regions and use it to evaluate the impact of several road projects financed by the Cohesion Fund. Results are presented in the form of multipliers, which give the real income effects generated relative to the direct impact of a project. The main advantage of this approach for project evaluation is the detailed microeconomic structure incorporated into the analysis. The main disadvantage is that this structure is imposed and not subject to econometric testing. Their analysis shows that sometimes a project in a single region can have strong welfare effects rippling through numerous regions. This is the case of the completion of the M-40 ring road around Madrid: because it acts as a link for some of the main Spanish motorways it has strong spillover effects throughout the regions in Spain and Portugal. Some other projects in a single region can instead have very localised effects. This is the case of the Tagus crossing in
Lisboa: because it improves mainly local transport costs, its effects are small outside the Lisboa region. Even projects connecting different regions can have relatively concentrated effects. For instance, the Rías Baixas motorway, connecting Galicia with the Spanish Meseta, creates gains concentrated mainly in Galicia. Combes and Lafourcade (1999) study the locational effects of changes in transport costs in France, using extremely detailed transport data. They find that reductions in transport costs between 1978 and 1993 have increased the concentration of employment in France.

6. Conclusions

Despite large regional policy expenditures, regional inequalities in Europe have not narrowed substantially over the last two decades, and by some measures have even widened. Income differences across States have fallen, but inequalities between regions within each State have risen. European States have developed increasingly different production structures. And European regions have also become increasingly polarised in terms of their unemployment rates.

This paper has argued that recent theories of location can help us explain these trends, and also reconsider the role of regional policies in such an environment. Firms producing in locations with relatively many firms face stronger competition in the local product and factor markets. This tends to make activities dispersed in space. However, the combination of increasing returns to scale and trade costs encourages firms to locate close to large markets, which in turn are those with relatively many firms. This creates pecuniary externalities which favour the agglomeration of economic activities.

Reductions in trade or transport costs, by affecting the balance between dispersion and agglomeration forces can decisively affect the spatial location of economic activities. For high trade costs, the need to supply markets locally encourages firms to locate in different regions. For intermediate values of trade costs, the incentives for self-sufficiency weaken. Pecuniary externalities then take over, and firms and workers cluster together. However, the price of local factors and the availability of goods tend to increase wherever agglomeration takes place. If this is the case and there is enough mobility, as trade costs continue to fall, rising factor prices simply give an additional kick to agglomeration by inducing immigration. On the other hand, if there is little mobility, for very low trade costs it may be firms that relocate in response to wage differentials. However, the combination of minimal interregional migration with institutional constraints on interregional (but not international) wage differentials can lead to a rise in income inequalities between regions within each country at the same time as inequalities between countries fall. Further, if agglomeration does not get reflected in wage differences, it may show in unemployment rate differences. Since clusters of activity may extend across borders, this can result in clusters of high and low unemployment extending across regions and even across countries. Firms tend to co-locate with firms to which they are more closely related and to avoid congestion by moving away from firms to which they are not. This also promotes increasing specialisation in a way which is not just be driven by traditional comparative advantage considerations.

Whether there is too much or too little agglomeration in the absence of regional policy interventions is not clear. The fact that firms and workers move without taking into account the possible
losses for those left behind implies there may be too much agglomeration. On the other hand, since when firms and workers move they do not fully take into account the benefits they bring for other firms and their impact on aggregate growth, there may be too little agglomeration. Thus, there is no general indication of the direction in which governments should push with regional policies when seeking efficiency. Even in terms of equity, the direction of policy is not obvious. Policies that increase agglomeration may nevertheless make those that remain in poorer regions better off by increasing production efficiency and the rate of growth.

Despite these ambiguities, European regional policies have the explicit aim of reducing regional inequalities. One of the main instruments for this is the improvement of transport infrastructure. However, it is not obvious that lower transport costs facilitate convergence. Roads and rail tracks can be used to travel both ways. A better connection between two regions with different development levels not only gives firms in a less developed region better access to the inputs and markets of more developed regions. It also makes it easier for firms in richer regions to supply poorer regions at a distance, and can thus harm the industrialisation prospects of less developed areas.

New economic geography models not only point out this potential ambiguity in the impact of lower transport costs on less developed regions, they also tell us that the overall effect depends on certain aspects of the economic environment (such as mobility and wage rigidities) and on characteristics of the projects. On this respect, the Trans-European Transport Network will give much of the EU better access to the main activity centres. However, the gap in relative accessibility between core and peripheral areas is likely to increase as a result of the new infrastructure, which reinforces the position of core regions as transport hubs. The emphasis on high-speed rail links is also likely to favour the main nodes of the network, and is unlikely to promote the development of new activity centres in minor nodes or in locations in between nodes.

References


Duranton, Gilles and Diego Puga. 2001. From sectoral to functional urban specialisation. URL http://dpuga.economics.utoronto.ca/. Processed, University of Toronto.


