I. Introduction

Regional inequalities are an inescapable fact of Europe’s economic geography. Close to one fifth of Europeans live in regions eligible under Objective 1 of the Structural Funds, the principal regional policy instrument of the European Union. The criterion for receiving this aid is a per capita income 75 percent lower than the European average. If such a criterion were applied to the United States, only two states (Mississippi and West Virginia), representing less than 2 percent of the American population, would qualify. The disparities in the average unemployment rate are also very wide in Europe, while these are practically nonexistent in the United States. They represent disparities not only among countries but also among the regions within each country: for the five “big” countries (Germany, France, Italy, Great Britain, Spain), the unemployment rate in the most affected region is at least seven points higher than in the least affected one.

As a reaction to these disparities, the EU has been devoting an increasing share of its budget to regional policies. The Structural Funds and the Cohesion Fund today represent over one third of the community budget. These transfers are of considerable macroeconomic significance: over the period 1994-1991 they represented 3.5 percent, 3.3 percent, 2.4 percent and 1.5 percent of GDP for Greece, Portugal, Ireland and Spain respectively. For the period 2000-2006, their impact will be less, in particular for Ireland. In terms of investment, they play a decisive role: respectively for those four countries, 15 percent, 14 percent, 10 percent and 6 percent of total investment is financed out of community resources.

Such intervention has had contrasting effects. Since 1989, there has certainly been convergence among the European countries, Ireland being the most successful and the most spectacular example. However, convergence among the regions within each country has not come about, indeed in most countries regional disparities have increased.

II. Regional convergence and divergence in Europe

A comparison between the economic geography of the EU and of the United States is instructive on more than one count. From the standpoint of regional income gaps, we have already noted that these were much wider in Europe than in the United States. But this larger disparity does not reflect a greater spatial concentration of economic activities: on the contrary, in Europe close to half of industrial employment is concentrated in 27 regions representing 17 percent of the continent’s territory and 45 percent of its population. In the United States, half of industrial employment is concentrated in 14 states representing only 13 percent of the entire country and 21 percent of its population. Midelfart-Knarrvik et al. show that, even taking into account the fact that industry is less concentrated in Europe than in the United States, most of the sectors are more geographically dispersed in Europe than in the United States. However, over the past 20 years the European countries have become increasingly specialized, and the structures of their industries have become more and more different. The empirical studies performed by Brülhart (1998), Brülhart and Torstensson (1996) and Amiti (1998) strongly suggest that a process of spatial concentration is under way in Europe today, going back in particular to the eighties, when good progress was made with trade integration.
Why has the greater spatial concentration of economic activities in the United States not engendered a larger income gap among the states? The first reason is the marked mobility of economic agents in the United States, a phenomenon that does not exist in Europe. In the United States, mobility among states is not only much greater than among the European countries themselves but also among the regions within a given European country. This strong American mobility explains why the phenomenon of spatial concentration of economic activities in the United States has not been accompanied by a process of per capita income divergence among the American states. The fact is, when workers follow mobile capital (physical or human) from regions in decline to regions experiencing growth, the problem of spatial equity becomes much less acute. Although certain regions are being emptied of their economic activities, per capita income does not diverge among the different regions. The emigration of workers from declining regions to growth regions makes it possible to reduce competition among workers in the former and to increase it in the latter. Such migrations are thus the principal force for adjusting regional inequalities in the United States. This certainly explains why, in the United States, the question of regional or spatial planning policies has never become such a significant issue as it is in Europe. To the contrary, it is because the workers (often the most disadvantaged) are also the least mobile in Europe that economic geography has taken on a political dimension only on this side of the Atlantic.

Another phenomenon noted in the introduction is the different way inequalities have developed among the European countries and among the countries’ own internal regions. Table 1 below illustrates the development of those disparities measured by the standard per capita GDP deviation for the NUTS2 regions. Only two countries, France and Germany, showed a lessening of inequalities over the period. In three countries, Belgium, Austria and Portugal, they remained somewhat constant. In the seven other countries for which regional data are available, inequalities increased. The two last lines of the table also show that while inequalities among countries diminished, those among the countries’ own internal regions on average increased.

Map A1, attached to this text, shows GDP growth rates per region and gives a clear picture: it is easy to recognize the areas that have “triggered” their country’s growth. For Spain, Catalonia and Madrid, for Portugal, the Lisbon region, for England the southeast, for Sweden Stockholm, for Finland Helsinki, and for the Netherlands the Amsterdam region. More detailed studies have shown (Duro, 2001) that up to the mid-eighties income inequalities among member States represented half of the inequalities among the European regions, and inequalities among regions within each State represented the other half. Since then, inequalities among States have diminished by 25 percent, but regional inequalities within the States have increased by 10 percent. As a result, the majority of regional inequalities in Europe are explained by inequalities within the countries. Thus, Europe is experiencing a process of convergence among countries at the same time as a process of divergence among the countries’ own regions: all of the convergence among the regions in Europe at the European level is thus explained by the convergence among countries.
Table 1: Regional disparities in per capita GDP within the member States, 1994-98

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>25.9</td>
<td>25.3</td>
<td>25.7</td>
<td>25.7</td>
<td>25.7</td>
</tr>
<tr>
<td>Germany</td>
<td>31.3</td>
<td>26.7</td>
<td>26.7</td>
<td>26.5</td>
<td>26.8</td>
</tr>
<tr>
<td>Greece</td>
<td>7.8</td>
<td>10.4</td>
<td>10.2</td>
<td>10.1</td>
<td>10.2</td>
</tr>
<tr>
<td>Spain</td>
<td>15.9</td>
<td>17.1</td>
<td>17.7</td>
<td>18.4</td>
<td>19.1</td>
</tr>
<tr>
<td>France</td>
<td>30.8</td>
<td>28.2</td>
<td>28.2</td>
<td>27.0</td>
<td>26.5</td>
</tr>
<tr>
<td>Italy</td>
<td>25.5</td>
<td>28.6</td>
<td>28.7</td>
<td>27.8</td>
<td>27.6</td>
</tr>
<tr>
<td>Netherlands</td>
<td>10.8</td>
<td>13.4</td>
<td>14.3</td>
<td>15.4</td>
<td>15.8</td>
</tr>
<tr>
<td>Austria</td>
<td>28.1</td>
<td>30.8</td>
<td>30.2</td>
<td>29.2</td>
<td>27.8</td>
</tr>
<tr>
<td>Portugal</td>
<td>13.8</td>
<td>13.5</td>
<td>13.3</td>
<td>14.0</td>
<td>14.2</td>
</tr>
<tr>
<td>Finland</td>
<td>17.1</td>
<td>18.3</td>
<td>21.2</td>
<td>22.0</td>
<td>24.6</td>
</tr>
<tr>
<td>Sweden</td>
<td>11.0</td>
<td>13.1</td>
<td>14.0</td>
<td>16.2</td>
<td>17.1</td>
</tr>
<tr>
<td>GB</td>
<td>18.3</td>
<td>31.4</td>
<td>31.7</td>
<td>33.4</td>
<td>33.9</td>
</tr>
<tr>
<td>EU15 (by member State)</td>
<td>12.7</td>
<td>12.5</td>
<td>11.9</td>
<td>11.5</td>
<td>11.2</td>
</tr>
<tr>
<td>EU15 (within member States)</td>
<td>23.0</td>
<td>24.5</td>
<td>24.7</td>
<td>24.8</td>
<td>25.0</td>
</tr>
</tbody>
</table>


A similar development in spatial polarization may be described for unemployment. Overman and Puga (2002) show that since the mid-eighties, regions starting out with a low or high unemployment rate have not shown much change in their relative situation. Regions with intermediate unemployment rates, on the other hand, have developed toward extremes. The authors interpret this result as an effect of the spatial polarization of economic activities due to economic integration. They show that the fate of the regions in terms of unemployment is linked much more closely to the results of the neighboring regions (whether or not they belong to the same country) than to that of the respective country itself.

II. The new economic geography

The introduction of economies of scale and of transaction costs may explain why regions with no obvious comparative advantage in certain activities can become centers of production of those activities. A model of the underlying mechanisms was introduced by Krugman (1991), who was at the origin of the so-called “new economic geography.” The central finding of this literature is that the diminution of transaction costs on trade may engender a concentration of economic activities in certain regions that have better access to the large markets even if they do not have the lowest production costs. This spatial concentration is advantageous because of the existence of economies of scale conducive to limiting production locations, and it is made possible by commercial integration which, while reducing transaction costs, does not oblige enterprises to be located close to all their consumers.

The interaction of economies of scale and transaction costs may be understood on the basis of a numerical example. Let us assume that an industry can locate in three regions: the Ruhr, a rich and central region with high wages and hence high labor costs, Catalonia, a middle-income region close to the large European markets, and Andalusia, a peripheral region with low wages and hence low labor costs. Economies of scale play a major role in the sense that unit production costs increase with the number of locations. Let us assume that the firm can produce in the three regions, in two of them, or in only one.
The choice of location is simply a minimization of the sum of production and transaction costs. The numerical example in Table 2 assumes that (1) production costs are higher in the Ruhr than in Catalonia, and higher in Catalonia than in Andalusia; (2) it is less expensive to concentrate production in one location because of the economies of scale; (3) ease of market access means that production in three locations minimizes transaction costs and that Andalusia is farther from the large markets than Catalonia, and Catalonia than the Ruhr.

<table>
<thead>
<tr>
<th>Table 2:</th>
<th>Production costs</th>
<th>Transaction costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Production in 3 locations</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Production Ruhr + Catalonia</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>Production Ruhr + Andalusia</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Production Ruhr</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>Production Catalonia</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Production Andalusia</td>
<td>11</td>
<td>16</td>
</tr>
</tbody>
</table>

What happens with the integration process, i.e. in this numerical example, what happens when the transaction costs are cut by half? At first, it is easy to see that when transaction costs are high, it is profitable to produce in the three regions in order to minimize transaction costs. We thus have the situation of a geographically dispersed industry since companies want to be located close to all their markets, both big and small.

When transaction costs are reduced, for example because of commercial and monetary integration or the integration of transport infrastructures, the firm may exploit the economies of scale and produce only in a single location. Which region is going to benefit from this concentration? In this example, we see that it is not, however, the region with the lowest costs that benefits from relocation. Catalonia has higher costs than Andalusia but has better access to the large markets, not only its own but also those within the core of Europe.

Thus in this example integration has a convergence effect, since Spain benefits from the relocation within its territory of activities that were at first partially located in a richer country. However, in Spain itself, activities are relocating from the poor and peripheral region to the wealthy region. Thus we have the phenomenon of local divergence.

It is clear, however, that if transaction costs were lower (for example, if they were close to zero) the region with the lowest costs would benefit from relocation. In theory, we have here a bell-shaped relationship between spatial concentration and transaction costs. This means that the poor and peripheral regions will attract enterprises for which the regional costs are relatively higher than the transaction costs, and for which the economies of scale are relatively small. On the other hand, the “core” regions will attract firms for which economies of scale and transaction costs are of prime importance. The study by Forslid, Haaland and Midelfart Knarvik (2002) shows that in Europe the location of industries seems to follow this pattern and that the relationship between concentration and commercial integration follows a different bell curve for each industry. Concentration is particularly strong when the firms involved are ones that belong to sectors with increasing returns, which seems to provide empirical confirmation of the studies on the “new economic geography.”
Recent literature entitled “new economic geography” exploits this interaction between economies of scale and transaction costs, but goes further, analyzing the cumulative phenomena that can come about. These can culminate in agglomerations based on worker migration (“American” model) or on relocation of the enterprises themselves (“European” model).

Let us take the example of the second agglomeration mode. An industry setting up in a region with few other businesses will need to import most of its inputs from other regions and will also have to sell the majority of its production in regions other than the one in which it is located. From that standpoint, it will be at a disadvantage compared to firms in other regions with high levels of industrial concentration. This will apply in particular to industries with strong economies of scale, namely those where market size is decisive for their profits. These are in general firms from sectors where the fixed cost (for example, the amount assigned to Research and Development before the start of the production phase) is high. On the other hand, businesses locating in a low concentration region will experience less competition in terms of attracting workers (and hence lower wage costs).

We therefore see that industries with strong economies of scale, belonging to sectors with a high profit margin, will automatically be attracted by those regions where other enterprises (both customers and suppliers) are already installed. In such regions a cumulative process is likely to be established, since the more a given region has been able to attract enterprises of this type, the more it will attract others. On the other hand, the regions from which those businesses have fled will find it increasingly difficult to attract other ones. Such regions will be able to attract firms whose wage cost represents a large share of total costs, in sectors that are highly competitive and thus posting low profit margins. But this cumulative process will mean a high level of specialization in regions with a heavy concentration of industries with high profit margins and wages as compared to regions that are able to attract only firms with low profit margins and wages. The former will specialize in high technology goods that do not suffer much competition from the opening up to trade of the low-wage countries. On the other hand, the latter will be directly affected by the process of trade globalization. This process of spatial concentration is not linear, given that the drop in transaction costs may have no impact at first. It is only when those costs have reached a critical level that the cumulative process takes shape. Once that process has started (once the former spatial equilibrium has been destabilized), it becomes self-maintaining and self-reinforcing.

III. Regional policies - tradeoff between equity and efficiency

Equity is one of the traditional motivations of regional or spatial planning policies. Certain economic agents, be they workers or consumers, are not mobile and are therefore condemned to live in poor or declining regions from which the mobile factors (capital and highly skilled workers) have departed. Because of the lower labor demand in such regions, real wages will adjust downward, or, if real wages do not adjust because of labor market rigidities, unemployment will increase. In both cases, the welfare of the inhabitants will deteriorate. As consumers, those agents will also see their welfare deteriorate since certain goods and services will no longer be produced locally (the businesses have left for more wealthy regions). In certain cases, in particular for certain services, the transaction cost will become so high that they can no longer be consumed by those agents. Thus the diversity of consumable goods and services in the poor region will decline. Moreover, the most mobile agents are in general those with the highest level of human capital (education, experience etc.). Such agents, thanks to the possession of “positive externalities” in the form of localized social interactions, have a positive impact on the productivity and thus on the real wages of other workers. By leaving a region in decline, the most productive workers thus also have a negative impact on the productivity of the remaining workers, i.e. those who are the most disadvantaged. There is thus an absence of market coordination, given that when certain agents decide on their location, they do not take into account the effect of their choice on the other agents. From that standpoint, there is a real market failure, with the consequent increase in inequalities.
that is specific to the spatial dimension of the economy and may thus serve as motivation for public intervention.

There are several ways to analyze the impact of the agglomeration phenomenon on the least mobile agents. The first would be to refuse to see it as a problem of equity but to interpret it as coming from a specific market failure. This would find its origin exclusively in the lack of mobility of the most disadvantaged agents, something we have already noted as characteristic of the European countries relative to the United States.

In Europe, promoting the spatial mobility of workers is not considered a solution to the problems of regional inequality. This is legitimate, but only partially, since because of cultural and sociological obstacles, there will always be a substantial fringe of workers who will be harmed by geographic inequalities. The vision of regions empty both of inhabitants and of economic activities (such as the Dakotas in the United States) is unacceptable in Europe.

A further motivation for public intervention at the regional level, put forward by the Commission, is that of efficiency. It sees in geographic disequilibria “an underutilization of economic and social potentials and an inability to take advantage of opportunities that could be beneficial to the Union as a whole.”

This motivation is much less clear than the equity based motivation. If the phenomena of spatial concentration are explained by the existence of economies of scale, this means that the spatial agglomeration is at the origin of economic gains. This will be the case if firms can benefit from the proximity of other enterprises in the same sector to diminish their costs (transport costs or fixed costs). This will also be the case if such concentration makes it possible to increase the firms’ productivity through localized spillover effects, i.e. if the firms can receive transfers of knowledge from other neighboring businesses. These localized spillovers have been documented in numerous studies (see, for example, Jaffe et al. 1993). The example of Silicon Valley shows the advantage a country can obtain from a very heavy spatial concentration of activities with positive technological externalities. The stronger spatial concentration of innovation-based activities in relation to production activities thus has an economic rationale, and the benefits of this spatial concentration go beyond private gains.

The objective of policies promoting a greater dispersal of economic activities is based on the assumption that the economic geography produced by market forces alone is too concentrated. However, the efficiency argument may demand more or less spatial concentration: on the one hand the economic gains of spatial agglomeration, and on the other the effects of congestion (pollution, for example). The fact that in Europe the convergence of countries is accompanied by national divergence makes one think that the former type of argument, efficiency gains with spatial concentration, has pride of place. In this case, a tradeoff between equity and spatial efficiency appears inevitable (see Martin 1999a, b). The graph below shows the positive correlation between per capita income growth and growth in regional disparities over the five-year period 1994-1998, and suggests the existence of such a tradeoff.
IV. Difficulty of assessing regional policies

One of the principal lessons from the studies on the effects of public policies on economic geography (see, for example, Puga 2002 and Baldwin et al. 2002) is that because of the cumulative, nonlinear processes under way it is very difficult to evaluate and anticipate the effect of public policies intended to offset tendencies toward concentration. Thus, regional policies which, for example, finance transport infrastructure, have effects in terms of both supply and demand. The demand effects are principally of a short term, typically Keynesian nature: the construction of a highway increases the local demand for goods and labor and hence increases the incomes available in the region. In turn, this increase generates expenditure on local goods (in particular non-exchangeable services), etc. In actuality, macroeconomic studies, such as those of the Commission (1999) have found a positive effect on regional growth over the short term, i.e. over a five- or six-year period of regional policies. But those macroeconomic studies look at the macroeconomic effect at the country level only, not the regional level. This is a problem, since we have seen that convergence exists among the European countries but not among regions in the same country. Nothing is said about regional policies benefiting the poorest regions of countries converging at a global level, but it is well known that the poor and rich regions of those countries did not converge among themselves over the same period. The pertinent study level is the region, not the country.

Other empirical studies have not been very encouraging for regional policies. This is particularly the case of Boldrin and Canova (2001) who find no effect of regional policies on regional convergence. Midelfart-Knarvik and Overman (2002) find that regional policies have encouraged R&D-intensive industries to locate in countries and regions that have low endowments of skilled labor. From the standpoint of efficiency, it is not obvious whether this is a great achievement.

More basically, it is essential to study the effects of these long-term policies, that is to say the effects on supply. Those effects, particularly as regards the location choices of industries, are more complex and may even be the exact opposite of the effects on short-term demand. This may be the case, for example, of transport infrastructure, which has received preferential treatment within the framework of European regional policies.
If we look at the case of industries with economies of scale, a policy to open up peripheral regions may have a paradoxical effect. By reducing transaction costs for inter-regional trade, such a policy may encourage firms to exploit their economies of scale by concentrating production in a single location. For firms whose wage cost is relatively low, this will mean concentrating production in the wealthy region, even if it means exporting part of that production to the poor region at low cost, thanks to the new transport infrastructures (see Martin and Rogers, 1995, and Martin 1999). The numerical example given in Table 1 may be reinterpreted as an experiment to set up transport infrastructure among the three regions, and it is easy to see that it culminated in the concentration of activities in one single region.

The recent study by Combes and Lafourcade (2000), suggesting that the drop in transportation costs in France over the past 20 years has resulted in an increased concentration of industries with strong economies of scale, shows that this result is not only theoretical. The example of Italy is also interesting: the construction of highways between north and south to open up the south does not appear to have had the expected effect, and has perhaps even accentuated the decline of the south (see Faini, 1983).

Of course, this outcome is not a widely generalized one. Martin and Rogers (1995) show that a transportation infrastructure that reduces transaction costs within a disadvantaged region will have a positive effect on that region, since it makes it possible to increase the effective size of the market in those regions. The public infrastructure that has enabled the northern region of France to come close (in terms of transportation costs) to the most wealthy European regions is at least partially at the root of that region’s renewed growth. In this case, transportation infrastructure has had the effect of making this a “core” region, and has not simply served to open up a peripheral region. Thus, industries have been able to concentrate their production in a region that has grown close to the large markets and also benefits from relatively low wages or at least a labor market that is very favorable to businesses. This example shows that the same public infrastructure will have very different effects from one region to another.

While part of the phenomenon of concentration and regional inequality is due to processes of localized “technological spillovers,” i.e. to the fact that enterprises located in the same region benefit, through social interaction, from a higher level of technological performance, the recommendations for public policies are very different. It is no longer a matter of reducing transaction costs on trade in goods among regions but of reducing transaction costs on the trading of “ideas.” This involves a change in priorities: rather than financing highways (with positive Keynesian effects), it will be necessary to promote technological convergence among regions, which involves public programs for telecommunications, the Internet, training of human capital, policies aimed at increasing the productivity of poor regions and facilitating the transportation of people rather than goods. Martin (1999b) argues that such policies make it possible to achieve gains in both regional efficiency and regional equity, in contrast to the regional policies financing goods transportation infrastructure or transfers to the poor regions.

Another difficulty with regional policies is that the goal of reducing regional inequalities is often confused with that of reducing inequalities among individuals. The fact is that regional policies consisting of subsidizing industries to enable them to relocate in disadvantaged regions may have exactly the opposite effect. If the capital is mobile, subsidizing the return on that capital in one region amounts to increasing its return in all regions. Regional policies for subsidizing capital, even if they succeed in reducing regional inequalities, can thus end up increasing inequalities among individuals, since the first beneficiaries may be those who hold the capital.

V. Conclusions
The low mobility of Europeans not only among countries, but, and more importantly, among regions of the same country, is an essential element making it possible to explain why the location of economic activities is a major political and social issue in Europe. On the one hand, it weakens the agglomeration process. Greater mobility would certainly involve the multiplication of regions empty both of inhabitants and businesses. On the other hand, it is this low mobility that is at the root of the human cost of spatial concentration, in particular in terms of unemployment, since the departure of companies does not involve the departure of workers. In the United States there is more or less a pattern where the workers passively follow the businesses, while in Europe, there is a demand for policies designed to encourage firms to follow the workers wherever they choose to live. Regional policies then appear as an alternative to human mobility: it is socially preferable to move activities to where the people are than to move people to activities.

In this paper, we have reviewed some of the challenges faced by regional policies. We have done this using some of the insights provided by the “new economic geography” literature. The main challenges we identified are the following: 1) a tradeoff may exist between spatial equity and spatial efficiency: the same economic forces, namely increasing returns, which explain spatial agglomeration are at the origins of economic gains. 2) policies that aim to reduce the peripherality of poor regions by reducing trade costs between poor and rich regions may lead economic activities to leave the poor regions. The long term supply effect of these policies may be the opposite of the short term demand one. In theory, policies that promote technological convergence seem more appropriate. 3) Reducing regional inequalities does not automatically reduce individual inequalities. The identification of these challenges is not meant to disqualify the legitimacy of regional policies in Europe but simply underlines simple rules of public economics that should apply to regional policies: it is essential to identify the market failures that legitimate public intervention, objectives (which inequalities to reduce) and possible tradeoffs between these objectives.

References:
Duro, J.A., 2001, Regional income inequalities in Europe: an updated measurement and some decomposition results. Instituto de Análisis Económico CSIC.


