The role and effectiveness of regional investment aid

The point of view of the academic literature

Written by Pierre-Philippe Combes and Tanguy van Ypersele

Competition

European Commission

The role and effectiveness of regional investment aid

The point of view of the academic literature

Report by Pierre-Philippe Combes and Tanguy van Ypersele

April, 2012

Luxembourg, 2013

Pierre-Philippe Combes:

Aix-Marseille School of Economics, Greqam, 2, Rue de la Charite 13 236 Marseille cedex 2, France. pp-combes@univmed.fr

Tanguy van Ypersele:

Aix-Marseille School of Economics, Greqam, 2, Rue de la Charite 13 236 Marseille cedex 2, France. tanguy.vy@univmed.fr

We thank Susana Peralta for her help in the writing of this report. We also thank Xavier Boutin, Kai-Uwe Kuhn, Ralf Martin, Thierry Mayer, Paul Seabright, John Van Reenen and Vincent Verouden for helpful discussions. Finally, we thank Richard Crabtree for editorial assistance.

The views and opinions expressed in this report are those of the two authors, not those of the Aix-Marseille School of Economics nor of the European Commission.

Luxembourg: Publications Office of the European Union, 2013

ISBN 978-92-79-28198-3

doi: 10.2763/32058

Contents

Summary

The purpose of this report is to study, from both the theoretical and the empirical point of view, the extent to which regional policy can reduce the disparities in economic activity levels that arise between regions belonging to an integrated trade area as the European Union. We start in Chapter 1 by reviewing the theoretical mechanisms underlying the general economic geography scenario, relating the movements of goods and people to regional disparities. The possible emergence of a trade-off between spatial efficiency and equity is illustrated. We emphasise the redistributive aspects of agglomeration. There are almost always diverging interests with respect to agglomeration. If one group gains, another group loses. Using a utilitarian welfare criterion (focusing on aggregate welfare), we show that for a broad range of model parameters, the laissez faire-solution is efficient, including when it leads to agglomeration. Nevertheless, the laissez-faire equilibrium can under certain circumstances also lead to an over-agglomeration of economic activity. When some societal aversion for inequality is introduced, the case for over-agglomeration is strengthened.

A large body of the empirical literature assesses the magnitude and sources of possible gains from spatial concentration. This literature is presented in Chapter 2, dealing first with the studies on the determinants of productivity, then growth, firm location choices and innovation. A number of methodological concerns, not always properly addressed in these studies, are set out in detail. This literature draws a number of robust conclusions, including, for instance, the systematic positive effect, even when possible econometric biases are circumvented, of the density of economic activity (defined as the number of employees per square kilometer) on productivity, firm location choices and innovation. Similarly, the access to markets other than the local one has a positive impact on these variables. In addition to density and market potential, new investments by foreign firms appear to be attracted by the stock of past such investments. The impact of other variables such as specialisation, diversity or the size of firms is less clear-cut. It is found to be more country-, sector- and period-specific. Using data on patents, patent citations and innovations, the existence of spatial technological spillovers that decrease rapidly with distance is also empirically shown. Importantly, when estimated structurally, whether on wage equations or firms' location choices, the main economic geography models identified in the theoretical literature are not rejected as potential explanations of agglomeration.

There is a specific, though not very abundant, literature that endeavours to directly assess the impact of the regional policies implemented in certain European countries. We describe these studies in detail in Chapter 3. There is no evidence that public infrastructure, of transport in particular, affects local productivity or firms' locations choices. However, possible reverse causality has never been addressed in these studies. Tax differentials between regions have a significant impact on location choices (even when possible econometric biases are removed), but the magnitude of their impact on firms' location choices or employment created is small. The role of European funds remains to be further investigated: it is found to be statistically non-significant in general, but it has a small significant impact in one of the most accomplished studies. Finally, national programmes of regional assistance are found to have statistically significant effects but again of small magnitude, possibly affecting regional employment but re-allocating it towards less efficient firms or regions.

Chapter 4 in this report returns to theory and shows how economic analysis can shed further light on the role of regional policy in extended economic geography frameworks. Typically, it is argued that

even if the classical first-best theory predicts that when regions decide policies non-cooperatively they induce an inefficient allocation of resources, there is some reason to be slightly more optimistic, for the classical second-best proviso that introducing further distortions in a distorted world can actually be efficiency-enhancing.

The last chapter is more prospective and goes beyond the strict survey to give some examples on how, by relating the academic literature on both theoretical and empirical economic geography, one should be able to deliver some insightful recommendations on regional policies.

Lessons for regional policy from economic geography (theory)

The European Union has designed cohesion funds as a development tool used to help laggard regions to catch up with the rest of the Union. The intention was for the assistance to be temporary, some hysteresis being implicitly needed such that, once the policy removed, the regions would not come back to their initial status. The attractiveness of such a policy would be questionable if its effects were reversed when the aids stop. Economic geography is a convenient framework to analyse the geographical allocation of economic activities, and the possible effectiveness of regional aids. Indeed, agglomeration or concentration of economic activities are typically phenomena exhibiting hysteresis, which this literature emphasises. A change in the demand level of the region or of available public infrastructure, a decrease in investment cost or an increase in factor productivity may trigger agglomeration forces in a region and therefore have long lasting effects.

In Chapter 1 of the survey we discuss the desirability of triggering the clustering of economic activity. In an economic geography setting in which the concentration of economic activity is the result of a balance between, on the one hand, agglomeration economies (i.e. Marshallian externalities: knowledge spillovers, labour market risk pooling, and vertical linkages with suppliers or customers) and the exploitation of increasing returns to scale and, on the other hand, dispersion forces like trade costs, congestion (on land or labour markets) or market competition.

The real value-added of the recent economic geography models (from Krugman (1991b), say) is that they pinpoint the self-reinforcing nature of the agglomeration forces that induce multiple equilibria in an imperfect-competition setting with trade costs and increasing returns to scale.

The mechanisms at work in all economic geography models are quite similar and due to the fact that agglomeration forces are self-reinforcing. An increased concentration of economic activity may therefore trigger a snow-ball effect. An increase in local activity increases local demand, which makes the local production more efficient, which itself increases the attractiveness of the location for workers and/or firms, which in turn increases demand, and so on. What matters is that local demand for the final good is affected by the agglomeration of firms or workers. This may be because of the migration of consumers, as in Krugman (1991b), creating so-called forward or demand linkages, or because of the existence of intermediate inputs provided by the agglomerating firms themselves, as in Krugman and Venables (1995), so-called backward linkages.

Given the large potential number of market failures in such a setting, efficiency of the laissez-faire solution is far from ensured. Moreover, agglomeration also has redistributive effects; it generally increases the welfare of the households in the region of agglomeration at the expense of the households in the "periphery". Understanding the balance between those different effects helps assessing regional aid policy.

According to Ottaviano and Thisse (2002): "the conventional wisdom supports the view according to which the concentration of means with the most productive region is often the optimal strategy to

maximise global income". From this perspective, agglomeration should typically be an efficient outcome. The first question that arises in this context is the efficiency criterion used. The Pareto criterion (under which the allocation of resources is deemed optimal in the sense that no one can be made better off without making someone else worse off) is likely to be too weak to draw any conclusions. Indeed, as already mentioned, the redistributive impact of the agglomeration are significant. The literature is typically not very explicit with respect to the welfare criterion used in economic geography models. Most papers use aggregate income as their welfare criterion. A notable exception is Charlot et al. (2006), who use different welfare criteria ranging from Rawls (focusing on the outcome of the worst off) to Kaldor Hicks (focusing on the mere potential for Pareto improvements).

Ottaviano and Thisse (2002) frame the debate around an efficiency versus equity trade-off. They argue that most of the policy debate takes this trade-off as given. Taking aggregate real income as an efficiency criterion, they show that in the "core-periphery model", agglomeration can be the resulting market outcome even though dispersion would be more efficient. The level of trade costs is an important determinant in this regard. The authors show that, in a context of declining trade costs, agglomeration takes place too soon, i.e. when it is not yet efficient. In that case, the efficiency equity trade-off does not exist (countering agglomeration forces through regional policy is good for both efficiency and equity). The main reason is that, in this second-best framework[1], prices no longer reflect the social opportunity cost; when deciding to move from one region to the other workers neglect the impact of their decision on the labour and product markets of both the hosting and the origin regions. However, these authors also show that when trade costs are low enough, agglomeration does become efficient.

Pfluger and Sudekum (2008) analyse a core-periphery model with a housing market. The particular feature of this model is that for low trade costs, dispersion re-emerges as an equilibrium outcome when the housing market is tight enough. They show, like Ottaviano and Thisse (2002), that agglomeration occurs too soon. When dispersion re-emerges for low trade costs, it is also too soon, i.e. for a range of trade costs, dispersion is the market equilibrium when agglomeration would have been more efficient.

In a subsequent paper, Charlot et al. (2006) extend their analysis to a broader range of welfare criteria. They use a social welfare function with an aversion-to-inequality parameter. Here the outcome heavily depends on the degree of aversion to inequality. At one extreme, with perfect neutrality, the results of Ottaviano and Thisse (2002) hold. At the other extreme, when the Rawlsian principle is used, dispersion is always efficient.

The next important issue of theory the survey addresses regards the determination of optimal public policy in those models. Public intervention can have an impact in two different dimensions. It can influence the level of agglomeration by acting on dispersion and agglomeration forces. This means that, given the type of equilibrium, public policy can seek to improve efficiency or make the allocation more equitable. Local public good provision, production subsidy or infrastructure investment have an impact on density. Typically, an increased supply of local transport infrastructure, schools, health services, etc. in one region would alleviate dispersion forces and therefore increase agglomeration in that region. We also discuss how transport infrastructure policy may have a differentiated impact whether it impacts intra-regional trade costs or inter-regional ones. A public infrastructure aiming at decreasing trade costs

[1]The theory of the second best concerns what happens when one or more optimality conditions cannot be satisfied. If one optimality condition cannot be satisfied, it is possible that the next-best solution involves changing other variables away from the ones that are usually assumed to be optimal.

within a region is likely to attract economic activity into the ailing region while one decreasing trade costs with the other regions is likely to lead to a decrease in economic activity.

The second area involves selecting one particular equilibrium when more than one co-exist. Public policy can either change expectations about the equilibrium that prevails or it can seek to eliminate undesirable equilibria. The theoretical literature shows that infrastructure policies, local public good provision and tax incentives are likely to influence the agglomeration process. The important question is whether local governments would be able to decide efficiently about those policies. A priori, the answer is negative as those policies, via their influence on agglomeration, generate inter-regional spillovers. Therefore, the non-cooperative equilibrium of the game played by the regions is likely to lead to an inefficient allocation of resources. Nevertheless, as the world considered here is of a second-best type, it is not obvious that the distortions introduced by the non-cooperative setting of local policies do not counteract other distortions. We will discuss this issue in more detail in the penultimate part of this summary.

The gains from spatial concentration (empirics)

The largest part of the empirical literature in economic geography is devoted to the identification and quantification of the impact of the characteristics of regions that enhance regional productivity. Though regional aid programmes are not always directly assessed in this literature, once the regional characteristics that are the most beneficial for productivity are characterised, one can then infer how regional aid can affect, or not, these characteristics, which is what gives these policies an indirect impact on productivity.

The variable that has attracted most attention is the local density of economic activity. Researchers now consider as robust the fact that a doubling of economic density increases local productivity by around 2-3% (one has to remember that density gaps between regions are often as large as a factor of 10, the total productivity effect being therefore quite large, higher than 10% in this case). Therefore, the question that has to be assessed is the extent to which regional aid can induce an increase in density. It must not be forgotten, as the theoretical models outlined in the previous chapter show, that regional aid that would increase density in one location may decrease it in other locations, where the impact on productivity would be negative. The overall empirical assessment of regional aid must therefore evaluate their impact on all regions.

Density is only one example of the regional characteristics for which the effect has been empirically evaluated. It belongs to a broader group of determinants named "urbanisation economies" that include all overall characteristics of the regional economy. By contrast, "localisation economies" concern the role of those characteristics that are specific to the industry in which the firm, or the worker, operates. Urbanisation economies include the impact of market potential or of the industrial diversity of the regional economy for instance. Localisation economies include the role of specialisation (the share of the industry in the local economy) but also some characteristics of the local firms in this industry, as their average size or their skills intensity.

It is most often found that economies of density diffuse across borders, the density in neighbouring regions also exerting a positive impact on regional productivity. More generally, market potential, defined as the spatially discounted sum of all regions' density, is systematically found to positively influence regional productivity. Following Hanson (2005), some papers follow a more structural econometrics route that leads to the role of a market potential variable that is corrected by price effects. This variable, directly suggested by many recent economic geography models, is also found to contribute to local productivity.

More generally, structural estimations of economic geography models tend, for many countries, not to reject these models as an adequate explanation, at least partial, of regional disparities.

By contrast, no robust role is found for industrial diversity. Specialisation has more often a significant positive influence on productivity but it is not a major determinant of spatial disparities. Lastly, local workers with highest skills, within the industry or not, are also found to exert a positive externality on other workers' productivity.

It is worth noting that the econometric estimation of agglomeration economies necessarily raises endogeneity concerns, be they due to possible missing variables or reverse causality. As is well known, such endogeneity may result in biased econometric estimates for the parameters under consideration. Recent advances in economic geography in the way to address such endogeneity have been made, further enlightening the debate on the role of regional aid. First, from Ciccone and Hall (1996), researchers have tried to carefully tackle reverse causality issues. Namely, economic density can increase productivity through a number of local externalities but higher productivity in some locations also attracts firms and people there. This in turn increases density, and so on. This reverse causality biases the simple OLS (Ordinary Least Squares) estimate of the extent to which productivity is influenced by density (the elasticity of productivity with respect to density). As it does not take into account feedback effects, OLS will typically overestimate this effect considered in isolation (the direct causal effect). Feedback effects affect almost all regional variables introduced in the different model specifications. The causal impact of regional characteristics on productivity can be obtained by a technique called instrumenting. Ciccone and Hall (1996) propose as instruments the historical characteristics of the location (typically its density decades or even centuries before the observation date). Other exogenous features of the regions can be used, as the geological nature of soils (Combes et al. (2010)) or their access to past infrastructure networks. Once density is instrumented, if regional aid were to impact density, one can thus distinguish their direct impact on productivity, the 2-3% gain from doubling density mentioned above, from an indirect further gain due to the endogenous location choices of agents. By inducing higher productivity in some locations, regional aid increase the incentives to locate there, which increases density and thus further enhances productivity. This leads to a total effect that has been estimated around a fifth larger than the sole direct one.

The second recent advance made by this literature concerns the use of individual panel data (observations of different variables over time) and the possibility to control for individual fixed effects when the impact of density is estimated. Combes, Duranton and Gobillon (2008) show that when individual characteristics are not properly controlled for, the direct impact of density is again over-estimated, but now largely, by a factor of 2. This is due to the fact that more dense areas attract more able workers. Again, this delivers interesting policy implications for the role of regional aid. Typically, if regional aid does not impact the local skills composition of the area, the gain of doubling density is only about the 2-3% we have just mentioned. Now, if regional aids succeed in changing the local skills composition of the area such as matching the one of the twice more dense areas, the gain could be up to twice larger. In the same vein, Combes et al. (forthcoming) show that agglomeration effects are much stronger for more efficient and larger firms (they benefit most from agglomeration).

Glaeser et al. (1992) initially proposed to study the impact of regional characteristics on employment growth and not on productivity. A somewhat large literature followed this route. However this raises an identification issue important for the design of regional policies since one does not identify separately the

impact on the employment of existing firms in the area from the impact on the creation of new firms. Combes et al. (2004) propose to decompose the total regional employment growth in the growth of the size of existing firms (internal growth) and the growth of the local number of firms (external growth) and to study the local determinants of the two separately, embodied in a Vector Autoregressive model (VAR, a statistical model used to capture the linear interdependencies among multiple time series, which also allows dealing with endogeneity issues). The conclusion is that indeed these determinants differ, which implies that regional aids affecting the regional characteristics of the economy, specialisation say, would have a different impact on these two components. More generally, using time series estimation strategies allows the researchers to tackle endogeneity (through the General Method of Moments (GMM)). However, these methods may suffer from weak instrument problems and the fundamental interpretation issue faced in this literature remains. One cannot assess if a negative effect of a regional characteristic is due to a negative impact of this characteristic on productivity, which then translates into lower employment, or, on the contrary, to a positive productivity effect that allows firms to reduce the employment they use.

A pretty large literature studies the impact of the regional characteristics we have just mentioned on the location choices made by foreign firms. These papers appeal to discrete choice models, following McFadden (1974). Head and Mayer (2004) for instance estimate a logit structural specification derived from an economic geography model that assesses the role of regional characteristics (among others market potential, which is augmented by some local labour cost and local tax variables) on the location choices of Japanese affiliates in European regions. The positive impact of the size of the regional economy and of market potential is confirmed in this study as in many other ones. Past foreign presence in the region, within the industry concerned or overall, is also found to have a systematic positive effect on new location choices made by foreign firms. Unfortunately possible missing variables and reverse causality biases are also present in such a FDI context but it is almost never addressed by researchers. For instance, local wages are often found not to significantly affect location choices but it may be simply due to the fact that the skill level of the regional labour force is not correctly controlled for.

Lastly, a literature followed the research strategy that Jaffe (1989) opened on US data, on the role of regional characteristics on innovation, assessed for instance through patents. In Europe as well, location is shown to matter a lot for innovation. It directly increases with the proximity to R&D sources. Conversely it then vanishes rapidly over geographic space. As for the US, patent citations are also found to be significantly larger in the neighbourhood of the patent location.

Direct assessments of regional policies (empirics)

In this chapter, we first recall that properly assessing the role of any policy is difficult due to the fact that one can never fully compare the situation of a region that has benefited from an aid to the situation of this very same region if it would not have benefited. We specify possible solutions, mostly based on difference-in-difference estimates and instrumentation. Unfortunately, we show that many studies attempt to assess the role of regional policy without considering this, and that very few seriously do it.

Transport infrastructure is not found to exert any role on productivity or on firms' location choices. This is, however, the typical example where, first, a strong reverse causality bias may be present due to the fact that regions targeted by new infrastructure are not randomly chosen. These are either backward regions that governments want to help or developing regions that need infrastructure investments. As a result, even after the infrastructure development they might remain less developed even if in a better shate than without. Second, economic geography models show that transport infrastructure is not only

a local input but also a means to connect different regions, facilitating trade between them. Improving infrastructure in a region can clearly be simultaneously beneficial, or detrimental, to other regions. Martin and Rogers (1995) emphasise that, if improving local transport infrastructure mainly benefits the region where it is improved, improving inter-regional transport infrastructure can be beneficial for one region only, the larger or the smaller region depending on which side of the bell-shaped curve the economy is. From the empirical point of view, this suggests that one should distinguish intra- and inter-regional infrastructure in the specification of the model. Importantly, it also requires interacting the role of infrastructure with the distribution of economic activities, as embodied in a market potential function for instance. None of this is typically done in the literature.

A number of articles assess the role of regional tax levels on firms' location choices. It is found to be negative but possible reverse causality (regions where more firms locate have more flexibility to decrease tax rates and can, more generally, have different characteristics from those who can only set higher tax rates), is typically not considered. Only two studies, Rathelot and Sillard (2008) and Duranton et al. (2011), seriously tackle such concerns by appealing to a similar empirical strategy for the UK and France, respectively. They restrict the sample to pairs of firms located very close to each other (less than 2 kilometres) but in different regions. The two firms constituting each pair are such that they face different regional tax rates but similar economic conditions in terms of market access, characteristics of the labour force, ie a number of variables that can be observed and introduced in the specification. They should be also similar in terms of unobservable variables. The strategy is further improved by an instrumentation of tax rates.

Rathelot and Sillard (2008) find that higher local taxes reduce firm creation but the effect is weak, such that increasing the tax rate differential by one percentage point increases the probability of a firm setting up in the lower taxed municipality by around one percentage point. Duranton et al. (2011) study not only firms' creation but also employment growth. They find that local taxation of non-residential property has a sizeable negative impact on employment growth, but no effect on entry. Importantly they show that methodologies that do not address the three problems of individual heterogeneity, unobserved time-varying location-specific effects and endogeneity of local taxation, give substantively different results.

As for the role of regional tax rates, there are some studies introducing variables controlling for the fact that firms or regions do benefit from European regional funds. In most cases the impact of these funds is found to be weak or non significant, even in the two studies (Dall'erba and Le Gallo (2008) and Mohl and Hagen (2010)) that instrument these policies (but not always in a completely satisfactory way).

Becker et al. (2010) choose another route and design an interesting regression discontinuity setting based on the fact that eligibility to the EU Objective 1 funds is confined to regions with a GDP per capita below the threshold of 75% of the EU average. The authors select a sample of Nuts 2 regions that are just below and above this threshold. They complement this research strategy by performing their estimation at the Nuts 3 level. This increases both the variability of the outcome and of the probability to benefit from the policy, and the comparability of the treated and non-treated samples. Indeed, some Nuts 3 regions within the treated Nuts 2 regions (thus below the 75% threshold) may well be themselves above this threshold, and therefore richer than other Nuts 3 regions present in the sample with a GDP below the threshold but not eligible because belonging to Nuts 2 regions above the threshold. Finally, because a time dimension is available in the data set, a difference-in-difference estimate is computed, which controls for regional characteristics.

It is found that Objective 1 funds do have an impact on regional growth. The programme participation exerts a differential impact on GDP per capita growth of about 1.8 percentage points within the programming period. With respect to employment, a statistically significant, but smaller, positive effect of about 0.5 percentage points is found. The authors propose a back-of-the envelope calculation that suggests that the funds spent on Objective 1 have a return which is about 20% higher than their costs. These conclusions are interesting but remain to be confirmed since, unfortunately, Becker et al. (2010) do not instrument the policy itself. Moreover, the main limit of regression discontinuity designs remains, that, strictly speaking, results are valid only for those observations in the neighbourhood of the discontinuity.

Finally, five studies using individual data, two for the UK (Devereux et al. (2007) and Criscuolo et al. (2012)) and three for France (Martin et al. (2011a), Mayer et al. (2011), and Briant et al. (2011)) assess the impact of some regional assistance (providing grants (for the UK) and tax exemption (for France) to selected firms) that have been implemented there.

Martin et al. (2011a) first show that the policy clearly targets firms located in disadvantaged regions and operating in declining industries, which underlines the importance to control for such a non-random selection of the firms entering the programme. The impact of the policy is studied on total factor productivity (TFP), employment, exports, and company survival rates; the estimation strategy mainly consists in a difference-in-difference approach with is extended to further robustness checks. The authors find no significant effect of French local productive systems on TFP, employment and exports, nor does it affect the survival rate of firms.

According to Mayer et al. (2011), the policy does have an impact on the probability that establishments locate in targeted areas. Importantly, they show that the impact of the policy is stronger for targeted areas that are initially less distressed and for sectors in which relocation costs are lower. Moreover, the supported areas tend to attract smaller firms, and not the larger ones. However, the analysis of the spatial pattern of the effect reveals that the policy does not create economic activity per se but rather operates as a firm relocation device within municipalities, inducing existing establishments, or new establishments to (re)locate in the supported part of the municipality. These results are broadly consistent with the findings of Briant et al. (2011) for the same French assistance programme. Briant et al. (2011) find that only a small positive average impact on firms and job creation rates, one that is strongly heterogeneous across targeted neighbourhoods. They show that the geographical characteristics of the neighbourhoods account for part of this heterogeneity. Spatial isolation, which accounts for urban severance and transport access, makes the programme less efficient.

Devereux et al. (2007) complement such estimation strategies by also instrumenting the grant received by firms and by simultaneously assessing the role of other agglomeration effects. They find that grants do have a statistically significant effect on average in attracting plants to specific geographic areas. However, the marginal effect is very low, implying that an increase in the expected grant of £100,000 is associated with a 1% increase in the probability of location, which evolves for instance from 1% to 1.01%. Importantly, firms are found to be less responsive to government subsidies in areas where agglomeration effects in the industry concerned are weak to begin with. Including the interaction term between the grant and the local number of plants in the industry, the estimated average marginal effect of the expected grant becomes three times higher. But the interaction term indicates that as the local size of the industry rises, the marginal effect of the expected grant does so also. An increase of 10 plants increases the marginal

effect of an increase of £100,000 in the expected grant on the probability of location by 6.7%. Therefore higher grant offers are needed to attract greenfield entrants to locations where industry agglomeration or natural resource benefits are weaker. This suggests that subsidies are less effective in influencing firms' location decisions in the face of countervailing co-location benefits. To address the fact that the sample of firms receiving grants is pretty small and that variables other than regional policy are endogenous, Wren and Jones (2011) use data aggregated by location and instrument all explanatory variables. There is a cost to that approach as a lot is lost in terms of the precision of the information used and the individual controls considered. Acknowledging this limit, they appeal to GMM estimation to instrument all variables. They estimate a significantly positive and concave effect of the grants. Each £25 million of grant changes the regional location of about six inward FDI projects. On average, projects have 150 jobs and each job diverted costs £27,500. Among other local characteristics that impact location choices, past FDI is found to have a strong positive effect. The impact of other variables is less intuitive, such as the negative impact of population and the positive ones of unskilled workers and of wages, or the absence of effect of the distance to major cities and of the skilled population.

Criscuolo et al. (2012) also evaluate the impact of the regional assistance programme in the UK but on employment and investment. As Devereux et al. (2007) they work at the individual firm level. They mix a difference-in-difference approach completed by matching with a simultaneous instrumentation of the policy implemented. They find that there is a large and significant average effect of the UK regional assistance programme on employment, investment and the probability of exit. These effects are seriously underestimated if endogeneity is ignored, as the participants in the programme appear to be weaker firms who would otherwise perform badly given their observable and unobservable characteristics. Importantly, there appear to be conversely no additional effects on productivity after controlling for the investment effects. Since the proportion of employment in entrants as a whole is shown to fall in areas benefiting from the policy, this raises the possibility of negative aggregate productivity effects (less reallocation of market share towards more efficient firms as inefficient incumbents are protected by the aid programme). Another important result is that the program has an effect on firms that is differentiated according to their size. Only small firms appear to be affected by the programme. On the other hand, they estimate that the cost per job of the program is only $6,300 suggesting that investment subsidies can be cost effective.

Regional aids, competition policy and fiscal competition (theory)

Regional aid from the European Union is an explicit exception to the general ban on state aid. We survey the literature analysing the potential impact of competition on state aid and more generally of fiscal competition.

In classic tax competition models, it is difficult to build the case for welfare improving state aid. The assumptions behind the classic settings (constant returns to scale, perfectly competitive markets, costless trade) are, however, quite restrictive. Once one abandons them in favour of more realistic setups, there is scope for efficiency-enhancing fiscal competition – this simply stems from the usual "theory of the second best" proviso that introducing further distortions in a distorted world may actually be desirable. Indeed, in imperfectly competitive settings, fiscal competition may have a corrective role. This happens if (i) the countries can design and commit to a tax schedule which is contingent on the full set of admissible strategies, (ii) firm subsidies are used to decrease the inefficiently high share of firms hosted by the core region, or (iii) to overcome inefficient locational lock-in. Put simply, when the subsidies are aimed

at attracting mobile firms, they are bound to allow the firms to internalise the externalities that their mobility imposes on the remaining economic agents. This is no longer the case when a state grants aid to its firms to avoid the exit of the most inefficient firms, following market integration. In such a case, the subsidies simply distort an efficient market mechanism.

The regional competition to attract firms with productive public infrastructure seems to be wasteful in most contexts. Infrastructure provision is increased over and above the efficient level in terms of enhanced firm productivity, leading to a total provision cost which outweighs the productivity gains. Again, one should be careful about devising policy recommendations based on a symmetric framework. Firstly, the result may be different when the economy is in a core-periphery equilibrium: in this case, the periphery actually under-provides infrastructure since it free rides on the core's infrastructure, whose benefits it can enjoy through lower import prices. This mechanism arises irrespective of whether agglomeration economies stem from mobility of the production factor or from the clustering of intermediate good suppliers. Secondly, allowing public infrastructure to be horizontally differentiated (i.e., different types of infrastructure as, for instance, the legal system, the transportation network, ...) or vertically differentiated (i.e., same type of infrastructure but differing in quality) may be used by the regions as a strategic device to soften tax competition, eventually promoting fully efficient tax setting and factor allocation in some cases. In this context, infrastructure serves the purpose of allowing the regions to capture a greater fiscal surplus from mobile production factors. Empirically, it seems to be the case that public inputs are not good mechanisms to attract capital, at least when accompanied by the corresponding, budget-balanced, capital tax increase.

There are many ways in which countries can manipulate their policy choices as a means to soften tax competition. This survey identifies three of them: public input differentiation, public input underprovision, and discrimination of different tax bases. Policy-wise, this insight advises for caution when regulating other policy instruments, for one may unintentionally worsen the fiscal revenue dissipation of the tax competition equilibrium.

Finally, the existence of appropriately defined fiscal equalisation grants seems to mitigate the inefficiency of tax competition in a variety of settings, even when it is efficiency enhancing compared to the no-tax equilibrium. The optimistic results are tilted in favour of equalisation of the tax base, as opposed to equalisation of tax revenue. While such grants already exist in many federal countries, there is scope for improving them along the lines identified in the literature. Importantly, the implementation of EU-wide fiscal equalisation schemes is bound to generate non-negligible efficiency gains.

Implications for State aid

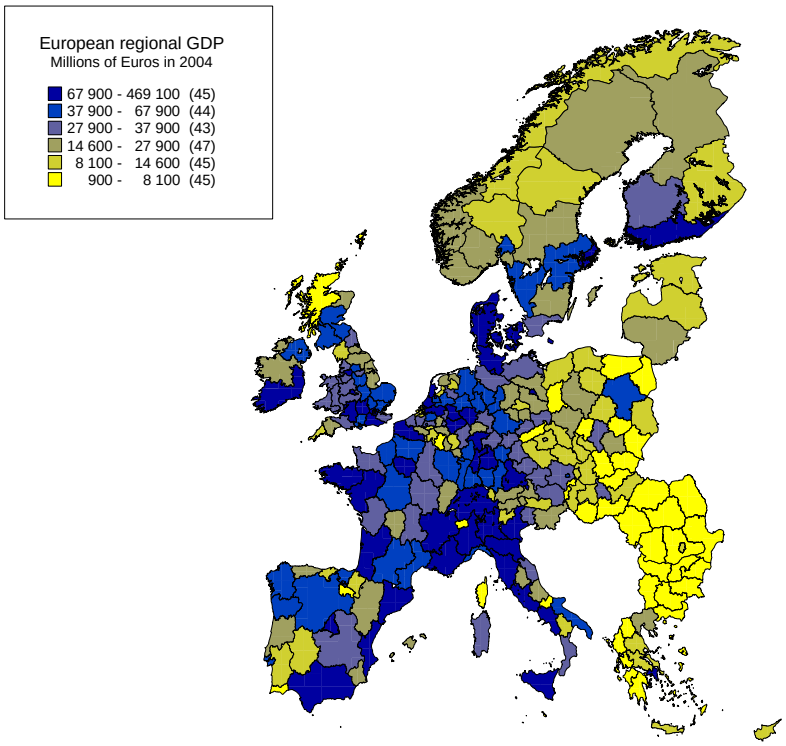
The policy implications of the literature's conclusions are interesting. If the EU's goal is to reduce disparities among regions via redistributive policies or state aid control policy, those policies have to be targeted to particular regions and particular firms. Devereux et al. (2007) hint at targeting regions where some agglomeration already takes place, Becker et al. (2010) to regions that are at the limit of being eligible, Mayer et al. (2011) to areas that are initially less distressed and to sectors in which relocation costs are lower, Briant et al. (2011) to regions that have a good market access. All this is fully consistent with the literature on fiscal competition and economic geography. Indeed, it shows that when all the activity is agglomerated out of a region, firms enjoy an agglomeration rent in the other location and therefore the subsidies needed to relocate industry there is large. Subsidies may also be targeted to particular type of firms. Both Criscuolo et al. (2012) and Mayer et al. (2011) show that small firms are

more responsive to subsidies. This is again in coherence with the literature showing that large firms can benefit more from economic density making them less willing to move to less dense regions.

Introduction

Regional disparities in Europe are large. This is a well-documented fact that policymakers bear in mind (see for instance European Parliament (2007)) and that we do not intend to fully document again here. Simple maps already illustrate the extent of disparities between European regions as a whole, i.e. in terms of total GDP (Figure 1), or between the individuals located in these regions, i.e. in terms of GDP per capita (Figure 2). In 2004, the region at the first decile had a GDP almost 15 times lower than the region at the last decile, and for GDP per capita the ratio was above 6. Another striking feature highlighted by the maps is the presence of some spatial auto-correlation in income and income per capita. Rich regions and rich individuals are surrounded by other rich regions and individuals, and conversely for poor ones. Finally, the contrast between a rich European core and a poor periphery also clearly appears.

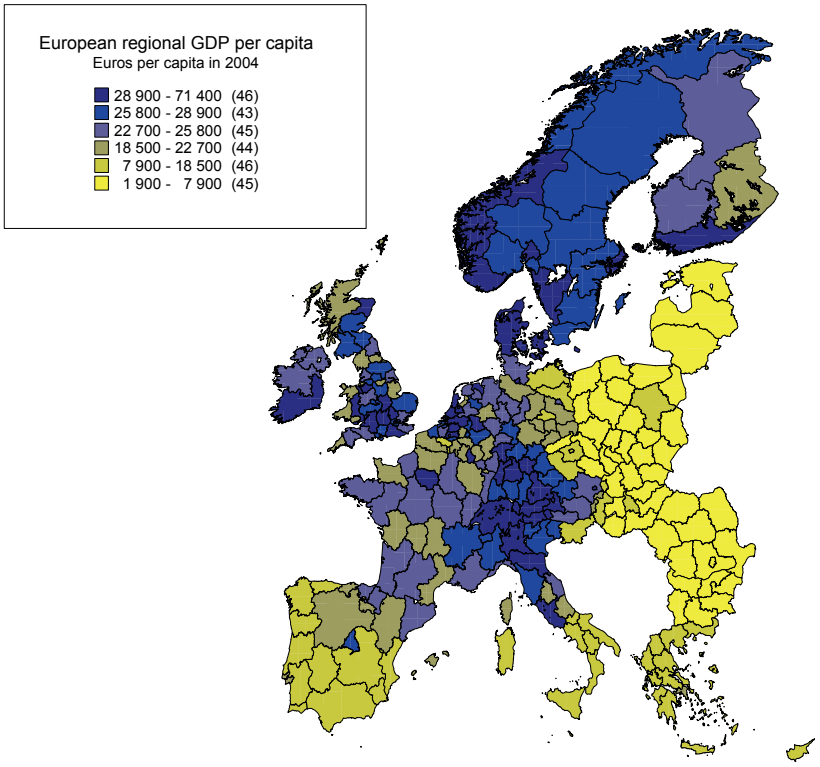
Figure 1: GDP of the NUT2 regions of the European Union in 2004 (number of regions in brackets).



Source: Combes, Mayer and Thisse (2008)

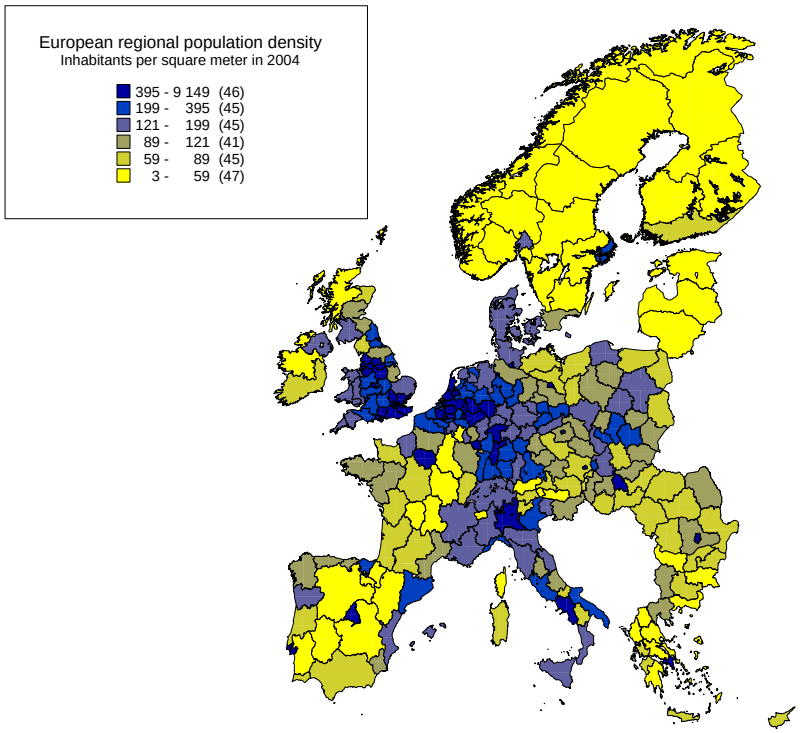
Economic geography experienced a new surge of interest in the 1990s, largely thanks to the work of Krugman (for which he won the Nobel prize in 2008). The main contribution of economic geography consists in relating regional income disparities to the spatial distribution of populations and presenting the corresponding concentrations. For instance, Figure 3 maps regional population density. It shares the same three main characteristics as GDP and GDP per capita. Disparities are large (density at the first decile is more than 15 times lower than density at the last decile), spatial auto-correlation is present, and core regions are denser than peripheral ones. Not surprisingly, the correlation of GDP and GDP per capita with population density is significantly positive, at 0.24 and 0.36 respectively.

Figure 2: GDP per capita of the NUT2 regions of the European Union in 2004 (number of regions in parentheses).



Source: Combes, Mayer and Thisse (2008)

Figure 3: Population density of the NUT2 regions of the European Union in 2004 (number of regions in brackets).



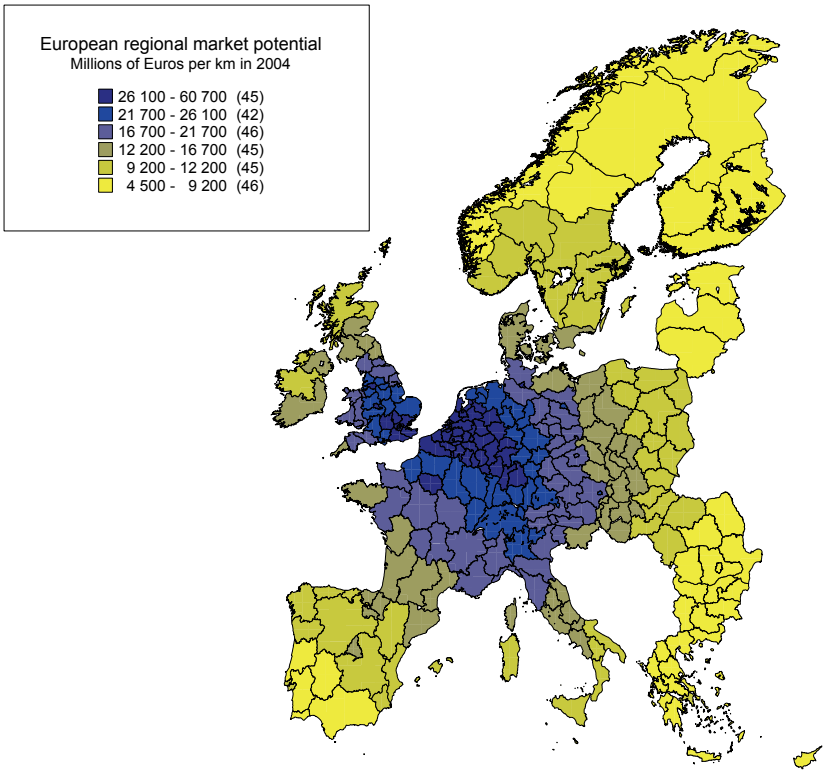
These regional characteristics are even more marked with another variable that was introduced by Harris (1954), namely market potential. For a given region, this consists in the sum of the GDP of all the other regions discounted by their distance from the region in question.

Economic geography argues that there is a circular economic causality between these variables - GDP and GDP per capita on the one hand, density and market potential on the other. It is believed that high density and high market potential, which are highly correlated (with a correlation coefficient at 0.44), cause a region to be richer in terms of both total GDP and GDP per capita. The main intuition is that dense regions are synonymous with large markets for the goods produced (which increases profitability when there are increasing returns to scale), with large markets for intermediate inputs and labour that make the goods cheaper, and with faster innovation and technology diffusion (which directly increase productivity). When goods are tradable, it is not only the region's density that matters, but also the access to other regions with whom one can buy and sell. Hence the role of market potential, when we make the reasonable assumption that trade costs are proportional to distance.

The other side of the coin is the reverse causality between income and density. Both labour and firms are mobile and choose where to locate partly on the basis of local economic conditions. Richer areas, which are the most profitable for firms and which provide the highest utility to households, therefore attract more activities, people and firms. This increases density and market potential disparities. When the two directions of causality combine, the result is a snowball effect: denser regions are more attractive, leading to an increase in density, which reinforces attractiveness, and so on.

The purpose of this report is to study, from both the theoretical and the empirical points of view, the

Figure 4: Market potential of the NUT2 regions of the European Union in 2004 (number of regions in brackets).



Source: Combes, Mayer and Thisse (2008)

extent to which regional policy can intervene in the above analysis. We start by reviewing the mechanisms underlying the general scenario, relating the movements of goods and people and regional disparities. The possible emergence of a trade-off between regional efficiency and equity is illustrated. We emphasise the redistributive aspects of agglomeration. There are almost always diverging interests with respect to agglomeration. If one group gains, another group loses. Using a utilitarian criterion, we show that the laissez-faire equilibrium can lead to an over-agglomeration of economic activity. Nevertheless, for a broad range of parameters, the laissez faire-solution is efficient, including when it leads to agglomeration. On the contrary, when some aversion for inequality is introduced, the case for over-agglomeration is strengthened.

A large body of academic literature assesses the magnitude and sources of possible gains from spatial concentration. This literature is presented in Chapter 2, dealing first with the studies on the determinants of productivity, then on growth, firm location choices and innovation. A number of methodological concerns, not always properly addressed in these studies, are detailed. This literature draws a number of robust conclusions, including, for instance, the systematic effect, even when possible econometric biases are circumvented, of the density of economic activity on productivity, firm location choices and innovation. Similarly, market potential has a systematic positive impact on these variables. In addition to density and market potential, new investments by foreign firms are attracted by the stock of past such investments. The impact of other variables such as specialisation, diversity or the size of firms is less clear-cut. It is found to be more country-, sector- and period-specific. Using data on patents, patent citations and innovations, the existence of spatial technological spillovers that decrease rapidly with distance is also proved. Importantly, when estimated structurally, whether on wage equations or firms' location choices, economic geography models are not rejected.

There is a specific, though not very abundant, literature that endeavours to directly assess the impact of the regional policies implemented in certain European countries. We describe these studies in detail in Chapter 3. There is no evidence that public infrastructure, of transport in particular, really affects local productivity or firms' locations choices. However, possible reverse causality has never been addressed in these studies. Tax differentials between regions have a significant impact on location choices even when possible biases are carefully removed, but the magnitude of their impact on firms' location choices or employment created is small. The role of European funds remains to be further investigated: it is found to be non-significant in general, but it has a small significant impact in one of the most accomplished studies. Finally, national programmes of regional assistance are found to have significant effects, but again of small magnitude, possibly affecting regional employment but re-allocating it towards less efficient firms or regions.

Chapter 4 in this report returns to theory and shows how economic analysis can shed further light on the role of regional policy in extended economic geography frameworks. Typically, it is argued that even if the classical first-best theory predicts that when regions decide policies non-cooperatively they induce an inefficient allocation of resources, there is some reason to be slightly more optimistic, for the classical second-best proviso that introducing further distortions in a distorted world can actually be efficiency-enhancing.

The last chapter is more prospective and goes beyond the strict survey to give some examples on how, by relating the academic literature on both theoretical and empirical economic geography, one should be able to deliver some insightful recommendations on regional policies.

Chapter 1

Lessons for regional policy from economic geography (theory)

The agglomeration of economic activity is a phenomenon that can be observed at different spatial scales. The finest scales consist in agglomeration patterns that can be detected even within cities, such as the concentration of the advertising agency industry at the block level in Manhattan studied by Arzaghi and Henderson (2008) or, at the city level, the famous Krugman (1992) example of the concentration of the US carpet industry in the city of Dalton (Georgia). The industrial districts in Emilia Romania in Italy are well-known cases of such highly specialised industrial clusters occurring at the regional level. At the other extreme, agglomeration forces are also at work at a continental scale, as with the US Manufacturing Belt (lying approximately within the parallelogram Green Bay-Saint Louis-Baltimore-Portland) or the European Hot Banana (a curved area stretching from Milan to London).

To understand the agglomeration of economic activities, without appealing to the standard trade theories based on comparative advantages that deliver clear cut and well understood conclusions, one must first consider the Starrett (1978) theorem: in a perfectly competitive economy, the presence of trade costs leads to a uniform distribution of economic activity, each location being autarkic. This implies that agglomeration can only result from a trade-off between increasing returns to scale in production - incompatible with the perfectly competitive economy - and transport costs. This is what governs the spatial distribution of economic activities. In the absence of transport costs, the number of production facilities would be determined only by the relationship between economies of scale and the size of the market. In the absence of economies of scale, each location would accommodate all existing economic activities, but on a very small scale. Increasing returns to scale and trade costs are therefore necessary basic ingredients of economic geography thinking.

Trade costs encompass all impediments to trade: transport costs, of course, but also trade barriers from policy or informational costs when one buys goods in non-local markets. Trade costs are fundamental in that they govern two important driving forces. The first is *market potential*. All else being equal, firms prefer to produce in the location with the best access to market because it minimises their trade costs. When trade costs are high, markets are segmented and production takes place close to the consumers. When trade costs fall, large markets become more attractive, as it is possible to produce locally and serve distant markets at a reasonable cost. The second force is the *market-crowding effect*. When firms concentrate in a region, it affects competition on all markets: it heightens competition in the host market

and reduces it in the others. The strength of this effect as a dispersion force depends on the level of trade costs. If there were no trade costs, the relocation of a firm would have no impact on competition, since whatever their location, firms have an equal access to all markets. With positive trade costs, a firm has easier access to its local market than to distant ones, and the location of the firm therefore affects competition.

Economies of scale (and their interaction with trade costs) are also fundamental. We shall consider different types of economies of scale. The most obvious are those that are internal to the firm. The larger the firm's scale of production, the lower its average cost of production. In the pre-industrial world, when distance-related costs were particularly high, production centres were numerous and their economic size was generally small. Firms were also small in size. During the industrial revolution, economies of scale became more prevalent and transport costs started to fall, leading to an increase both in the size of firms and in spatial concentration.

Economies of scale can be external to the firm. The firm's average cost of production decreases with the level of production taking place "close" to the firm. Clearly, these external economies of scale may be operating at different geographical scales depending on the industry and, once again, on the costs of interacting with other locations (which define what is close). External economies of scale may come from the local markets for inputs or goods. At the level of regions or large cities, a very wide range of services and inputs are offered to the firms, thereby increasing the productivity of other inputs (Hanson (1990)), including labour. Chapter 3 will present extensive empirical evidence on this. Moreover, in large cities, firms also find a wide range of workers' skills, while workers have access to a large number of differentiated job opportunities, enabling them to make the most of their skills and reduce their job search costs. The result is better matching between jobs and workers in labour markets which, simultaneously, tend to fragment and diversify (Hamilton et al. (2000)). In short, the division of labour becomes more fine-tuned as a result of the diversification and specialisation of tasks and this improves efficiency.

A similar phenomenon is at work in product markets. An increase in the range of varieties supplied is another driving force of regional development. As the fall in transport costs has gradually led to the disappearance of geographical monopolies, the resulting increase in competition has prompted firms to restore their profits through product differentiation (Irmen and Thisse (1998)). Consequently, firms in dense regions face more diversified intermediate inputs. As in local labour markets, the matching between the supply of intermediate inputs and firms' needs is improved, which has a positive impact on productivity.

Finally, external economies of scale may come from *communication externalities*, which are critical in a number of fields, including management, administration, research and finance. Knowledge, ideas and, above all, tacit information, can be considered as impure public goods that generate spillover effects from one firm or institution to another. The evidence of such effects will be described in detail in Chapter 3, but as an example, let us mention Jaffe et al. (1993), who compare the places where patents were taken out with those where they were cited. Having controlled for the impact of geographical concentration of a sector on the location of patent users, they show that the frequency of local citations of patents is systematically higher. Furthermore, this holds whatever the geographical scale under consideration (state or metropolitan area). Several studies confirm the "local" character of the exchange of ideas and innovation, at least in the first stages of the diffusion process. This phenomenon can be explained by the fact that the transmission of ideas that have not yet been formalised cannot take place in a standardised

way. The initial steps in the development of a new technology, say, require repeated contacts between the agents involved, to establish a common language, interpret individual pieces of information and bring them into the operational state. Such a process is facilitated by spatial proximity. Even in the age of telecommunications, "knowledge crosses corridors and streets more easily than oceans and continents". This last category of effects, which largely underpin endogenous growth theory, have been somewhat neglected by recent economic geography models, which focus more on local external economies of scale arising from market interactions, be they on labour, intermediate or final goods markets.

In short, the fall in transport costs and the progressive removal of trade barriers has enabled firms to agglomerate in a small number of locations or regions in order to benefit from local complementarities and externalities, while preserving the possibility of serving distant markets at low cost. We shall broadly call all the economies of scale coming from outside the firms "Marshallian externalities", as they correspond to external effects generated by other firms. The main objective of this chapter of the survey is to explain the forces emphasised by recent economic geography models and their implications for regional policy. We present the different workhorse models of the literature and then discuss their implications for efficiency and redistributive concerns.

1.1 Multiple-equilibrium, hysteresis and the scope for regional policies

The description of agglomeration forces dates back to the 19th century and the fundamental contribution of Marshall (1890). The real value-added of the recent economic geography models (from Krugman (1991b) say) is that they pinpoint the self-reinforcing nature of the agglomeration forces that induce multiple equilibria in a completely characterised imperfect-competition setting with trade costs and increasing returns to scale.

The mechanisms at work in all economic geography models are quite similar and rely on circular causality. Agglomeration forces are self-reinforcing, and an increased concentration of economic activity may therefore trigger a snow-ball effect. An increase in local activity increases local demand, which makes the local production more efficient, which itself increases the attractivity of the location for workers and/or firms, which in turn increases demand, and so on. What matters is that local demand for the final good is affected by the agglomeration of firms or workers. This may be because of the migration of consumers, as in Krugman (1991b), creating so-called *forward or demand linkages*, or because of the existence of intermediate inputs provided by the agglomerating firms themselves, as in Krugman and Venables (1995), so-called *backward linkages*.

1.1.1 The Core-Periphery model with migration: *Forward or demand linkages*

Krugman (1991b) illustrates in a parsimonious way circular causality and how it can generate multiple equilibria. The main force driving agglomeration in this model is the migration of workers. Increase the number of consumers in a region, it attracts more firms, which in turn decreases local consumption prices. This generates a new inflow of consumer-workers, which attracts even more firms, and so on. This is *forward linkage*.

The intuition of the model is easy to grasp, but its mathematical details are pretty complex. As a result, a number of results only received formal analytical proofs many years later Krugman (1991b) and

some are still only available in the form of numerical simulations.

There are two ex-ante identical regions (A and B), two sectors (manufacturing and agriculture) and two factors of production (agriculture workers and manufacturing workers). The factor endowment of the two regions is identical. Agricultural workers are assumed to be immobile between the two regions, while manufacturing workers are perfectly mobile between them, choosing to locate in the region that provides the higher utility level. The agricultural sector produces a homogeneous good under constant returns to scale using immobile workers that constitute the unique input. Perfect competition is assumed to take place in that sector, the good produced is freely traded and therefore has the same price in both regions. This price is used as the numéraire.

The manufacturing sector is characterised by Dixit-Stiglitz monopolistic competition. A large number of differentiated varieties is produced by firms using an increasing returns to scale technology that consists in a set-up cost plus a constant marginal cost of production. The mobile workers are used as the only input. All varieties are traded inter-regionally with, in this case, iceberg (proportional) trade costs.

In the Dixit-Stiglitz framework, manufactured goods are horizontally differentiated and consumers have a preference for diversity. The assumption of monopolistic competition means that firms supplying individual varieties act as monopolists for their own variety and neglect the impact of their behaviour on the aggregate price index. In this setting, consumers demand a composite manufacturing good which is made with all the differentiated varieties. For given prices of the differentiated goods, the larger the number of varieties, the lower the effective aggregate cost of consuming the composite good. This is because of the preference for diversity.

The first force is the presence of increasing returns to scale at the firm level. This obviously encourages firms to maximise their size by locating as close as possible to large markets. This is a clear agglomeration force. The second force at work is a dispersion force: the market-crowding effect. More firms in a given location means a smaller share of the demand for each firm. In the present model, as demand partially moves with workers, market-crowding effects operate mainly through the immobile workers' demand. In other words, the market-crowding effects are proportional to the size of the agricultural sector in the economy. Clearly, the higher the trade costs, the more segmented the regional markets, and therefore the stronger the dispersion force, since market-crowding differs more between the two regions. Note that the market-crowding effect can be expressed in terms of workers rather than firms. Less workers means lower local demand and therefore less firms and less competition, leading to higher prices, which in turn reduces the attractiveness of the region for the workers.

There is a third force that is linked to labour markets. More firms in a region means higher demand for workers, which increases the nominal wage, thereby making the region less attractive for firms but more attractive for workers. Conversely, more workers means a larger supply on local labour markets and therefore a downward adjustment of the equilibrium nominal wage, making the region less attractive for workers but more attractive for firms. These effects can act as a dispersion force when labour mobility is low, as it decreases the attractivity of the large region for the (more mobile) firms while increasing the attractivity of the small region of origin. It can be an agglomeration force when workers are very mobile, since the labour supply effect counter-balances the demand effect and nominal wages in the larger region increase less, and can even fall, as in the Krugman (1991b) model when trade costs are low and the regions are not too different in size.

We can now describe the circularity of agglomeration effects more precisely. When a larger number

of varieties is produced in a region, the price index of the manufacturing composite good is lower. This makes the location attractive to workers because they enjoy a higher standard of living. When workers move to the larger region they increase demand, causing the entry of more firms and an increase in the supply of local varieties. This further decreases the aggregate price index, and so on. The causality can only be broken if dispersion forces are strong enough: either when market-crowding effects are strong enough to prompt some firms to stay in the small region (i.e. when the number of immobile workers there is large) or when the excessive inflow of workers in the large region brings down the nominal wage too far, more than offsetting the effect of a lower price index. Note that the higher the trade cost, the greater the impact on the price index of the increased number of varieties locally produced. At the limit when trade costs are zero, the price index is unaffected by the number of varieties locally produced; it is only affected by the total number of varieties. Consequently, both agglomeration and dispersion forces are reinforced by an increase in trade costs.

The central question is how to determine when agglomeration forces dominate dispersion forces (in which case agglomeration should be the equilibrium outcome), and when the opposite occurs, (in which case dispersionshould be the equilibrium). An equilibrium is defined such that no worker has any incentive to move from one location to the other, either because they are indifferent between locations, obtaining the same utility level in both, or because they are in the location that provides the higher utility. At an interior equilibrium (i.e. an equilibrium with manufactured varieties produced in both regions and therefore with some mobile workers in both regions), the utility of workers has to be the same in both locations. Figure 1.1 shows, for different values of trade costs, the difference between the two possible locations, regions A and B, of the representative consumer's utility, as a function of the share λ of the mobile workers located in region A.

Because of the perfect symmetry between the regions, dispersion $\lambda = 1/2$ is always an equilibrium. When mobile workers are equally distributed between regions, the number of varieties produced in each region are the same, and so wages and prices are also the same in each region. However, this equilibrium may be unstable.

An equilibrium is stable when a worker who would have been forced to move to the other region is willing to come back to their region of origin because their indirect utility is higher there. (Equally, this can be expressed with respect to the movement of a firm). Here, we have two effects working in opposite directions. On the one hand, the labour market where nominal wages decrease as the labour supply increases makes the region less attractive for workers. On the other hand, it makes the manufacturing sector more profitable, leading to an increase in the number of varieties supplied. The local cost of consuming the aggregate manufacturing good therefore falls, thus generating an increase in real income.

When the first effect dominates the second, the dispersed symmetric equilibrium is stable, which is the case in Figure 1.1 for trade costs τ_1 or τ_2. It can be shown that this is the case when trade costs are above a threshold level τ_b. In this case, the difference in utility is decreasing in the neighbourhood of the symmetric equilibrium, which means that when one worker changes location he makes the indirect utility in the region he moves to lower than in the other region, hence his incentive to move back to the initial region.

A second type of equilibrium may prevail in which all mobile workers are located in only one of the regions. This is called the *agglomerated equilibrium*. This equilibrium exists if, when all firms and all workers are located in this large region, the workers get a higher utility there than in the other region.

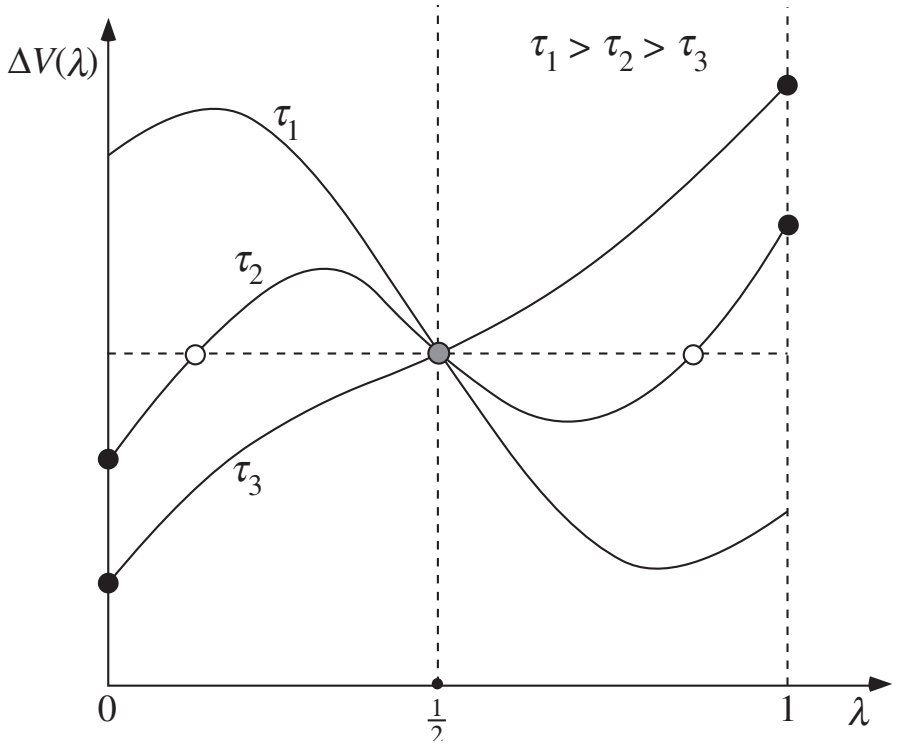


Figure 1.1: Migration dynamics for different values of trade costs

Source: Combes, Mayer and Thisse (2008)

One can show equivalently that firms do not have an incentive to produce in the periphery. Agglomeration is an equilibrium if a firm paying a wage that would induce the worker to move to the periphery is not able to break even (to make positive profits) there. The trade-off is the following. On the one hand, there is the market-crowding effect: the firm located in the small region is serving a market that is partially sheltered from competition. On the other hand, the wage paid to the worker has to compensate for the increased cost of living, since this worker has to import and pay trade costs on almost all the varieties he consumes. It is shown that the second effect only dominates the first when trade costs are below a threshold τ_s, as is the case in Figure 1.1 for τ_2 and τ_3. Interestingly, $\tau_b > \tau_s$ means that when trade costs are between those two values, both type of equilibrium coexist.

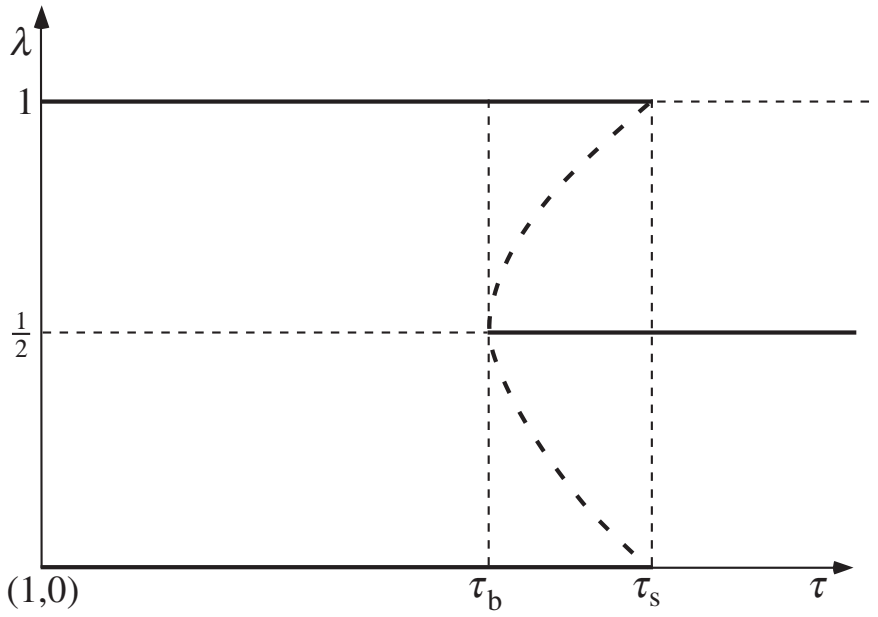


Figure 1.2: Set of equilibria in the core-periphery model

Source: Combes, Mayer and Thisse (2008)

Figure 1.2 summarises these findings and plots the allocation of the manufacturing sector as a function of trade costs. When trade costs are low, the only type of equilibrium is the core-periphery one, where one of the two regions hosts all the manufacturing sector. Note that both regions can potentially be the core or the periphery, as two such asymmetric equilibria exist. When trade costs are high, the only equilibrium is the dispersed one with perfect division of manufacturing between the two regions.

As we have just observed, for intermediate trade costs, both the dispersed and the core-periphery equilibria coexist. Selecting one type of equilibrium or the other has redistributive implications. Mobile workers have the choice of their location and therefore choose what is best for them. That is not the case for immobile workers, whose welfare is clearly affected. In this model, immobile workers would strictly prefer the core to the periphery in the agglomerated equilibrium.

Figure 1.3 plots the utility of an immobile worker in each of the locations as a function of trade costs. When trade costs are high, the economic activity is dispersed and immobile workers get the same utility in both locations. The lower the trade costs, the less costly it is to consume the imported varieties. The

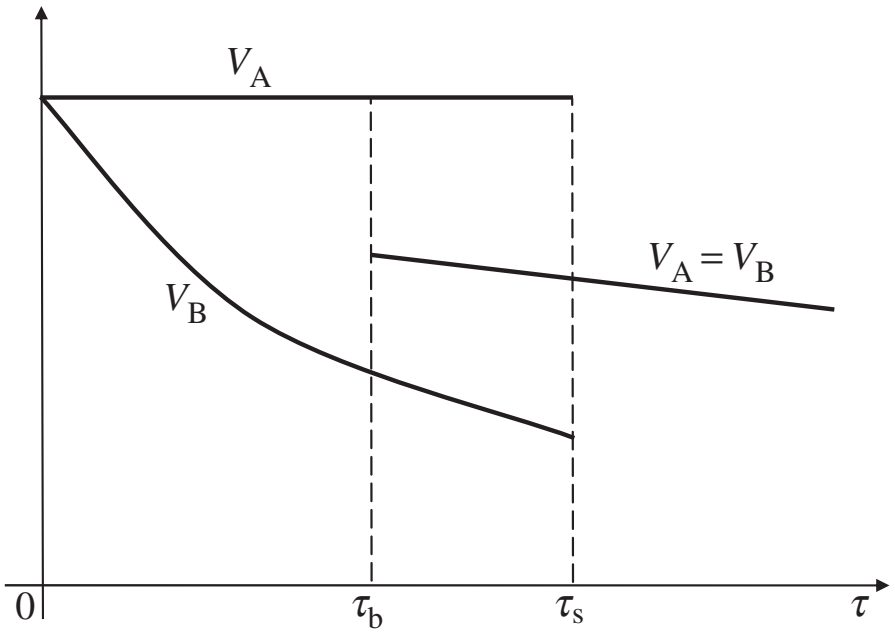


Figure 1.3: Indirect utility of an immobile worker in regions A and B

Source: Combes, Mayer and Thisse (2008)

utility of immobile workers increases as trade costs decrease.

When trade costs are intermediate, both the dispersed and the agglomerated equilibria exist.The utility of the immobile workers in the agglomerated region (say region A) is higher than it is in the other region. In the core, no variety has to be imported and therefore the utility of the workers is independent of trade costs. In the periphery, consumers have to import all the varieties of the manufactured good. The utility is therefore lower than in core, but also lower than the one they would obtain in the case of dispersion. Clearly, immobile workers in the periphery are positively affected by an decrease in trade costs, but they would prefer the symmetric equilibrium to be selected, which is possible since the two equilibria co-exist in this case. This is no longer the case when trade costs decrease further. When they become very low, the economy is necessarily in the asymmetric equilibrium where immobile workers experience a lower utility level than those in the core. Still, workers' utility converges across regions with the decline in trade costs.

One of the major shortcomings of this model is its analytical difficulty and the necessary use of simulations to obtain some of the results, even if Fujita et al. (1999) already present a number of contributions that prove analytically some of the results that Krugman (1991b) only illustrated by simulations. The literature has developed other models sharing most of the characteristics of Krugman (1991b) but more tractable analytically. They are known as the *Footloose Entrepreneur* models from Forslid and Ottaviano (2003). The difference from Krugman (1991b) is that immobile workers can be used in the production in both sectors. With the assumption of perfect tradability of the agriculture good between regions, this implies that the marginal cost of production is identical in the two regions. Mobile entrepreneurs are used only to set up firms. One entrepreneur is needed to start the production of one variety and he has to live where he produces. We still have forward linkage, as an additional entrepreneur increases both the

local demand for manufacturing goods and the number of locally supplied varieties. This forward linkage generates an agglomeration process that is self-reinforcing as in Krugman (1991b), and the conclusions in terms of the existence of symmetric and asymmetric equilibria and on how they depend on trade costs are similar to those drawn by Krugman (1991b).

Lastly, to illustrate how important migration is to the circular causality of agglomeration forces, we also briefly discuss what is call the *Footloose Capital* model. This is identical to the footloose entrepreneur model except that entrepreneurs (renamed capital) are assumed to set up firms in the most profitable region while spending their income in their region of origin. In other words, capital owners (and their demand) do not move with the capital. There is no forward linkage. Therefore, when regions are symmetric, the equilibrium is always stable and no agglomeration can take place. Agglomeration is induced by exogenous country asymmetries via the so-called *home market effect*. Because of increasing returns to scale, the larger region hosts a number of firms that is more than proportional to its size in terms of population. This can lead, for small trade costs, to full agglomeration of capital, and therefore of the manufacturing sector, in one sole region. Importantly, because of the absence of forward linkage, the circular causality of agglomeration is not present and the model does not feature multiple equilibria.

Krugman (1991b) has been extremely influential, because it shows how agglomeration can be the result of self-reinforcing agglomeration forces. Nevertheless, the main mechanism is based on the inter-regional mobility of workers, for which there is little evidence either between countries or even within them. Moreover, one of the predictions of the model is that nominal wages are lower in the centre than in the periphery, which appears to be counterfactual. This unattractive prediction is due to the lack of proper congestion effects. Higher wages have to be paid in Paris than in the provinces, for example, to compensate for higher housing costs. Lastly, some Marshallian agglomeration forces, such as the role of technological spillovers, of matching effects on the labour market or of intermediate inputs are absent from this model.

1.1.2 The Core-Periphery model without migration but with intermediate inputs: *Backward linkages*

Krugman and Venables (1995) overcome some of the previous shortcomings. The model proposed is close to Krugman (1991b). There are two regions, a manufacturing and an agricultural sector. The agricultural sector operates under constant returns to scale and perfect competition and does not incur any trade costs. The manufacturing sector is characterised by Dixit-Stiglitz monopolistic competition. On the demand side, agents have the same preference for variety. The novelty of Krugman and Venables (1995) is that there is only one type of labour that can be used in both agricultural and manufacturing sectors. Labour is therefore perfectly mobile between sectors, but it is now assumed to be inter-regionally immobile. This eliminates the forward linkage emphasised in Krugman (1991b).

Circular causality comes from the introduction of an intermediate input that is used in the production of the manufactured varieties. Both the fixed and variable costs of production result from a combination of the intermediate input and labour in a Cobb-Douglas function. The intermediate input is none other than the composite manufacturing good. As a result, an expansion of the manufacturing sector has two reinforcing effects: (1) it increases the local demand for the manufacturing good used as intermediate input, (2) it reduces the marginal cost of production of the local varieties, as it reduces the aggregate

price of the manufacturing good with the same preference-for-variety effect that shapes final demand. This is called *backward linkage*.

This linkage can lead to the expansion of the manufacturing sector and therefore a shrinking of the agricultural sector in one region, while the opposite occurs in the other region. Full specialisation, where agriculture completely disappears from one region can correspond to an equilibrium configuration. In this case, a new demand effect emerges, since in this case the wage in no longer tied to the wage in the agricultural sector, identical in the two regions. Increased manufacturing production now also increases the nominal wage, which increases the wage bill of the firms and reduces their profitability but also creates a forward linkage, not due to a larger population as in Krugman (1991b) but due to richer workers combined with higher local spending.

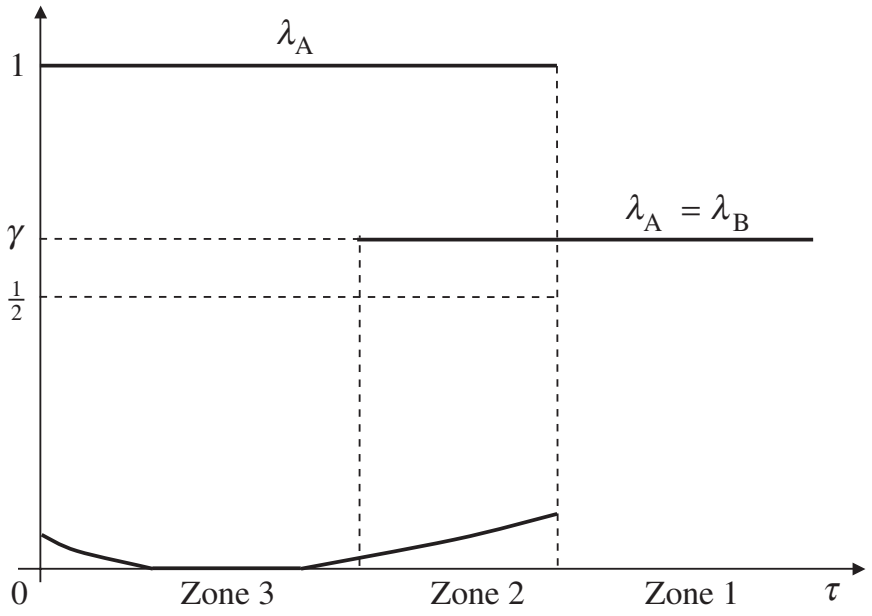


Figure 1.4: Manufacturing shares and trade costs for high final demand of the manufacturing good

Source: Combes, Mayer and Thisse (2008)

Not surprisingly, multiple equilibria can exist in this model. Figure 1.4 describes, for high final demand of the manufacturing good, the share of the labour force (λ) allocated to the manufacturing sector in each location as a function of trade costs.

When trade costs are high, there is dispersion in the sense of identical shares of manufacturing, numbers of varieties and price index in both regions.

When trade costs are low, core-periphery allocation of economic activity arises. The nice feature of this model is that it is possible for the periphery to keep some manufacturing production. The main reason for this is the inflated nominal wage in the centre, due to full specialisation of the core in manufacturing.

If the demand for the manufacturing good was lower, this feature would not emerge, because agglomeration would not lead to full specialisation of the core region in manufacturing. In that case, as in Krugman (1991b), the periphery is fully specialised in the agricultural sector. Figure 1.5 describes the impact of economic integration on the allocation of economic activity and on the utility of the workers.

Three phases can be distinguished.

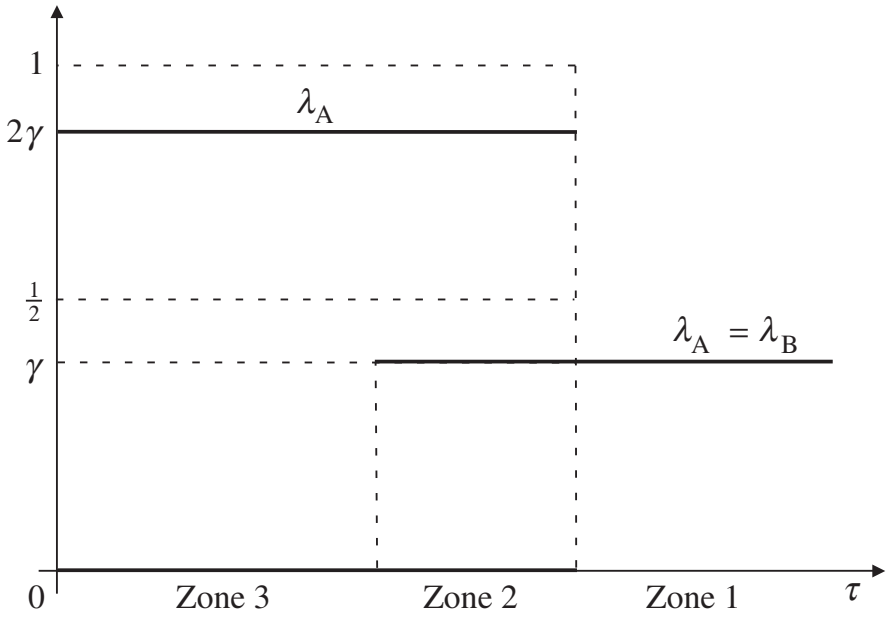


Figure 1.5: Manufacturing shares and trade cost for low final demand of the manufacturing good

Source: Combes, Mayer and Thisse (2008)

Phase 1: dispersion and harmonious development.

When trade costs are high, there is dispersion of economic activities. Regions are symmetric, hosting the same share of manufacturing, and the indirect utility of workers is the same in both regions. In this phase, a fall in trade costs benefits both regions; first, via its direct impact on the price index induced by the saving in trade costs, and second, indirectly via the increase in the equilibrium number of varieties. Since the manufactured good is an intermediate input, when the price index falls, the production cost of manufactured varieties also falls, thereby increasing profitability and attracting new entrants. This further decreases the price index, and so on.

Phase 2: agglomeration and divergence.

With intermediate trade costs, disparities emerge when the economy switches to the core-periphery equilibrium. The core region specialises (although not necessarily always fully) in the production of the manufactured good. This generates instantaneous gains for workers in the centre as the number of varieties that are locally supplied increases and so the price index directly decreases. A symmetric direct loss is incurred by the workers in the periphery.

There is also an indirect effect via the decreased cost of production, due to the impact of agglomeration on the intermediate input price and the gains related to backward linkage. Symmetrically, production costs increase in the periphery with the opposite consequences.

The aggregate effect of agglomeration on the number of varieties is positive because an increasing number of firms are better able to exploit increasing returns to scale. Further welfare gains may be secured in the core region when it fully specialises in the manufactured good. At that point, the nominal wage rises above the marginal productivity of labour in the agricultural sector.

Further decline in trade costs increases the relative number of varieties in favour of the core. This amplifies the effects described above and therefore increases the divergence between the regions. Nevertheless, the increase in nominal wage in the centre has a negative impact on its production costs and therefore reduces the production- cost gap between the core and the periphery during this phase. The increased number of varieties produced in the centre also increases competition, generating a market-crowding effect. These two effects are dominated for intermediate trade costs.

Phase 3: re-industrialisation of the periphery and convergence.

In this last phase, the dispersion effects due to high nominal wages and market-crowding dominate and trade integration therefore leads to re-industrialisation of the periphery. When trade costs are low, the gains obtained by not shipping goods are small, which weakens the local production-cost advantage. In this case, the nominal wage gap between the centre and the periphery makes production in the other region profitable. Then, further decline in trade costs increases the relative number of varieties in favour of the periphery, which in turns reduces the cost advantage in the core and also weakens the market size asymmetry between regions. It induces efficiency gains and increases the total number of varieties.

In this process, the core may gain or lose depending on the relative importance of the intermediate good in the production process. The periphery gains and the gap between the regions declines.

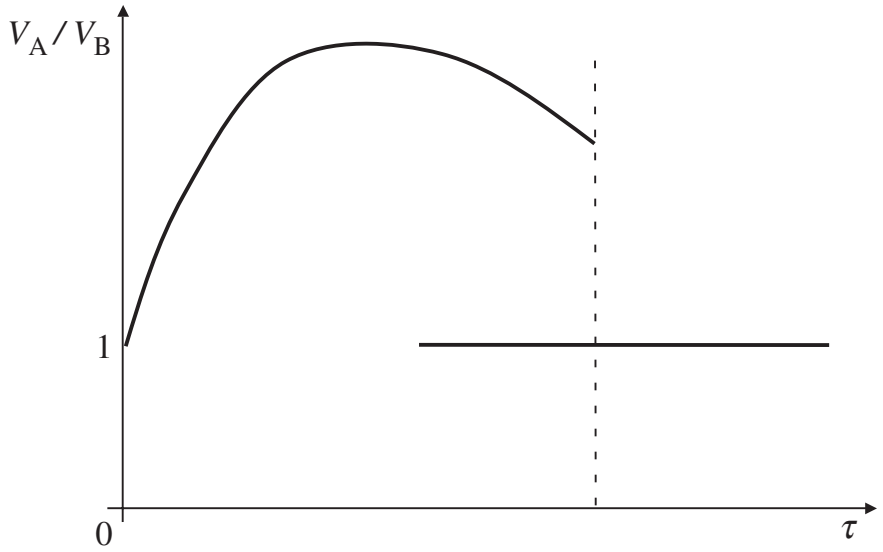


Figure 1.6: Welfare levels and trade costs: the bell-shaped curve

Source: Combes, Mayer and Thisse (2008)

As shown in figure 1.6, the relative utility of the workers in regions A and B follows a bell-shaped curve. With high trade costs, there is dispersion and utilities are identical. Both agglomeration and dispersion forces strengthen, but the former more strongly, and agglomeration occurs. At further integration levels, agglomeration and dispersion forces keep strengthening, but now more strongly for the latter and re-dispersion occurs, inducing a convergence of indirect utilities.

Krugman and Venables (1995) focus on two dispersion forces: market-crowding on the varieties market and labour market congestion. Subsequent literature (see for instance chapter 7 in Combes, Mayer and Thisse (2008)) has shown that other congestion costs, if strong enough, could lead to a comparable bell-

shaped relationship between regional utility differentials and trade costs. Land can act as a congestion force. If firms or mobile workers were to consume land or housing, the agglomeration of economic activity would increase its price and could therefore, at some point, also generate re-dispersion. It is interesting to note that if land entered the picture, the redistributive impact of agglomeration would change drastically. What matters is who uses or consumes the land.

Puga (1999), for instance, introduces land into the agricultural sector. In that case, any movement towards agglomeration causes an increase in nominal wages. The specialisation process moves workers from the agricultural sector to the manufacturing sector. This increases the amount of land per worker and therefore, on the one hand, drives up the marginal productivity of labour and wages, and on the other hand, depresses the marginal productivity and remuneration of land in the agricultural sector of the centre. The opposite effects are observed in the periphery. This acts as an additional dispersion force that becomes so strong that dispersion becomes an equilibrium when trade costs are low. In Puga (1999), the introduction of land amplifies the redistributive effects of agglomeration. A weird prediction of the model is that the price of land is lower in the centre than in the periphery, since its use is more intensive there.

On the contrary, if land is used by the manufacturing firms or by final consumers, agglomeration increases housing/land prices in the centre and reduces it in the periphery. Here again, land acts as a dispersion force, but the mechanism is quite different. For instance, in Pfluger and Sudekum (2008), land is used as housing by workers. Agglomeration implies a movement of workers from the periphery to the centre, inducing a similar shift in the demand for land. At least part of the agglomeration gains in the centre and disagglomeration losses in the periphery are capitalised in the land rent. Pfluger and Sudekum (2008) show that for low trade costs, economic activity re-disperses. Contrary to Puga (1999), in Pfluger and Sudekum (2008) the introduction of land attenuates the redistributive impact of agglomeration because part of its gains and losses are capitalised in the land.

Note that this sort of re-dispersion of economic activity for low trade costs can be sustained by other mechanisms. The literature has studied two of these in some detail. The first is based on the presence of trade costs in agriculture. In this case, a new effect of agglomeration is an increase in the agricultural good price in the core, since demand is greater there, while production is not. This encourages workers to move to the small region, all else being equal, as shown for instance by Picard and Zeng (2005). Another dispersion force may result from the idiosyncratic preferences workers have for some regions. Independently of the utility differential between regions, some workers prefer to live where their old friends are, in the region that has a similar climate to the one they are used to, etc. Tabuchi and Thisse (2002) and Murata (2003) show that this creates a dispersion force because even when the periphery de-industrialises, some people stay there, encouraging firms to locate close to them to benefit from the low market-crowding effect in this region.

1.1.3 Economic geography and growth

Growth can also act as an agglomeration force. Using a dynamic version of a model similar to Krugman and Venables (1995), Martin and Ottaviano (2001) introduce an R&D sector that is at the origin of new varieties. This sector generates a backward linkage of the type described in Krugman and Venables (1995) as it uses both labour and the composite manufacturing good in its process.

The cumulative agglomeration force operates as follows: more R&D activity in one region induces a higher local demand for the composite manufactured good, inducing a relocation of the production of some varieties there, which further reduces the price of the composite good, and so on. Assuming free trade in patenting, i.e. a variety discovered in one region can be produced in the other without extra cost, they show that as soon as the economy invests in R&D, the only stable equilibrium is one with full agglomeration of R&D activity and at least partial agglomeration of the manufacturing production. The impact on the utility of the workers in the periphery is ambiguous, as agglomeration fosters growth, and the gains linked to the increased number of varieties may therefore offset the losses due to the increased transport costs.

In an earlier paper, Martin and Ottaviano (1999) analyse the importance of local and global spillovers in R&D activities on the agglomeration process. They use a footloose capital model in which capital is the patent stock. They assume some Marshallian knowledge spillovers, since discovering a new variety reduces the cost of producing subsequent ones. These spillovers are local when only the local R&D cost is affected or global when the R&D cost is affected in both regions.

Not surprisingly, they show that when spillovers are global, there is no agglomeration of R&D and the growth rate is unaffected by geography, i.e. the localisation of economic activities. They also show that in equilibrium, because of the home market effect, there is a flow of FDI from the large to the small market, i.e. some varieties produced in the small market are owned by agents located in the large market.

When spillovers are local, the R&D sector agglomerates in the large market, and the rate of innovation is increasing with the concentration of firms there. Therefore, trade integration has a positive impact on growth.

The redistributive aspects of economic agglomeration are not as obvious as the ones in the static model. Agglomeration is not always detrimental to the immobile agents in the small market, as it fosters growth in both regions. Therefore, when spillovers are substantial, agglomeration increases welfare in both regions.

1.2 Too much or not enough agglomeration?

According to Ottaviano and Thisse (2002): "the conventional wisdom supports [] view according to which the concentration of means with the most productive region is often the optimal strategy to maximise global income". Agglomeration should be an efficient outcome. The first question that arises in this context is the efficiency criterion used. Pareto is likely to be too weak to draw any conclusions. Indeed, the redistributive impact of the agglomeration process described above is such that neither dispersion nor agglomeration are likely to Pareto-dominate. The literature is pretty poor with respect to the welfare analysis in economic geography models. Most papers use aggregate income as their welfare criterion. A notable exception is Charlot et al. (2006), who use different welfare criteria ranging from Rawls to Kaldor Hicks.

Ottaviano and Thisse (2002) frame the debate around an efficiency versus equity trade-off. They argue that most of the policy debate takes this trade-off as given. Taking aggregate real income as an efficiency criterion, they show that in the core-periphery model, agglomeration can be the market outcome even though dispersion would be more efficient. More precisely, they show that the level of trade costs at which agglomeration is efficient is lower than the level that triggers agglomeration. Somehow, agglomeration

takes place too soon when trade costs decline. In that case, the efficiency equity trade-off does not exist. The main reason is that, in this second-best framework, prices no longer reflect the social opportunity cost; when deciding to move from one region to the other workers neglect the impact of their decision on the labour and product markets of both the hosting and the origin regions. However, they also show that when trade costs are low enough, agglomeration is efficient. Pfluger and Sudekum (2008) analyse a core-periphery model with a housing market. The particular feature of this model is that for low trade costs, dispersion re-emerges as an equilibrium outcome. They show, like Ottaviano and Thisse (2002), that agglomeration occurs too soon, being non-efficient. They also show that when the housing market is tight enough, the model predicts that dispersion will re-emerge for low trade costs. In that case, dispersion also emerges too soon, i.e. for a range of trade costs, dispersion is the market equilibrium when agglomeration would have been more efficient.

In a subsequent paper, Charlot et al. (2006), the authors extend their analysis to a broader range of welfare criteria. They first use the Kaldor and Hicks criterion. This states that one equilibrium dominates another if the agents who prefer the first equilibrium are able to compensate those that prefer the second. This means that agglomeration is efficient if it is possible for the immobile workers and the mobile workers in the centre to transfer resources to the immobile workers in the periphery to make them indifferent between both types of equilibrium. Accordingly, dispersion is efficient if the immobile workers in the periphery are able, also by transfers, to make the other agents indifferent. Charlot et al. (2006) show that when trade costs are low, agglomeration is socially optimal. When trade costs are large, the Kaldor Hicks criterion cannot distinguish between the two types of equilibrium, as none of the winners are able to compensate the losers at the market price. They then turn to a social welfare function with an aversion-to-inequality parameter. Here the outcome heavily depends on the degree of aversion to inequality. At one extreme, with perfect neutrality, the results of Ottaviano and Thisse (2002) hold. At the other extreme, when the Rawlsian principle is used, dispersion is always efficient.

Ottaviano and van Ypersele (2005) show that in a footloose capital model, when trade costs are so high that no inter-regional trade occurs, or so low that full agglomeration takes place, the market leads to a second-best efficient (in utilitarian terms) allocation. The social planner would not be able to increase the aggregate real income by forcing the relocation of firms. However, for intermediate trade costs, the large region hosts too many firms and full agglomeration occurs too soon again. Baldwin et al. (2004) perform the same type of analysis for a broader range of asymmetries between countries. They show that these results need to be qualified because they rely on the large country having a higher per capita income. They show that if the large country has a lower per capita income, it will have too few firms.

All those papers seem to suggest that there is room for regional policy intervention. In what follows we will analyse the optimal provision of different local public policies. We leave to the last chapter the positive analysis of the non cooperative setting of those policies by regional public authorities.

1.3 Public intervention

Public intervention can have an impact in two different areas. It can influence the level of agglomeration by acting on dispersion and agglomeration forces. This means that, given the type of equilibrium, public policy can seek to improve efficiency or make the allocation more equitable. The second area involves selecting one particular equilibrium when more than one co-exist. Public policy can either change

expectations about the equilibrium that prevails or it can seek to eliminate undesirable equilibria.

In what follows, we survey different papers that analyse these questions. Some of these papers have a broader scope, in seeking to determine whether or not it is efficient to leave this type of policy tool to the regional government. In this section we concentrate on the impact of the first set of policies. The second set will be examined in Chapter 4.

1.3.1 Public intervention to change the level of agglomeration

Transport infrastructure policy

Martin and Rogers (1995) analyse infrastructure policy in a footloose capital model with two countries. In their paper, the aim of infrastructure policies is to reduce trade costs. It is assumed that trade is costly even within a country. Two types of infrastructure are considered. The first is intra-national and aims to reduce trade costs within a country; the second is international and aims to reduce trade costs between the two countries. By making trade easier within the country, intra-national infrastructure policies act as an expansion of the local market. Therefore, the increasing returns to scale sector tends to agglomerate in the country with the best local public infrastructure. This is akin to the home market effect. Martin and Rogers (1995) also show that a fall in international trade costs favours the region with the best local infrastructure. That is to say that an improvement in international infrastructure amplifies the concentration effect of the differential in domestic infrastructure. Their policy recommendation, if the objective of the policy-maker is to encourage international convergence, is therefore that the infrastructure policy should be biased in favour of the local infrastructure of the poorest countries.

Monfort and Nicolini (2000) carry out the same type of exercise in a more complex geographical setting. Whereas Martin and Rogers (1995) assume two countries and do not analyse possible agglomeration within countries, Monfort and Nicolini (2000) have two regions within each country. Labour is assumed perfectly mobile between regions of the same country but totally immobile between countries. We therefore have a centre-periphery model within each country. The authors analyse the impact of international trade integration on the agglomeration within countries. They show that international integration increases within-country regional disparities.

In a similar framework, Monfort and van Ypersele (2003) show that the sequence of infrastructure policy changes matters. For instance, it may be in a country's interest to first invest in local infrastructure in order to reduce regional trade costs and thereby favour agglomeration within the country. This agglomeration is shown to give a comparative advantage in the production of the manufactured good in international trade. It is also shown that agglomeration in one country makes agglomeration in the other country less likely. Behrens et al. (2007) add a global welfare analysis in a similar model. They analyse the impact of a unilateral reduction in regional transport costs from a situation of symmetry. They show that when regional transport costs are high, a unilateral reduction in transport costs is socially undesirable. It is only when transport costs are low that a unilateral deviation may be profitable. Moreover, they show that when agglomeration takes place in both countries, a unilateral reduction in transport costs is always socially desirable.

In a completely different setting, Martin (1999), using the growth model of Martin and Ottaviano (1999) discussed in 1.1.3, analyse different public policies aiming to promote convergence between North and South regions. He shows that an improvement in the South's intra-regional infrastructure induces a

relocation of production from the North (where R&D takes place) to the South. Because of a backward linkage in the R&D sector, this relocation reduces the rate of innovation and therefore also the growth rate. The competitive impact of the reduced innovation rate leads to an increase in the return on capital. This is because the higher the innovation rate, the fiercer the competition on the manufactured market, and hence the lower the return on capital. Therefore, the improved infrastructure in the South has a positive impact on the return on capital. As capital is assumed to be owned in a larger proportion by households located in the North, this policy leads to a divergence in inter-regional income. The net effect on real income is ambiguous: nominal income inequality increases but the aggregate price index decreases in the South compared to the North.

Martin (1999) also analyses the impact of an improvement in the inter-regional infrastructure. He shows that it induces the opposite effect to that caused by a decrease in intra-regional trade costs. It attracts more production to the North, increasing R&D, which in turn decreases the return on capital. Nominal wage disparities decrease but the impact on the price index is more complex. For the North and the South, the increased equilibrium number of varieties and the decrease in trade costs exert a downward pressure on the price index. A third effect operates through the relocation of production between the South and the North. The South imports more varieties; the North less. It is shown that the total impact is positive for both North and South and that the price index falls more in the North than in the South.

Local public goods

Egger and Falkinger (2006) and Fenge et al. (2009) analyse the provision of local public goods in a core-periphery model, with a special emphasis on non-cooperative provision. In this section we only discuss the impact of the provision of public goods on agglomeration. The results specifically related to the non-cooperative setting are surveyed in Chapter 4.

Both papers assume that the local public good is destined for firms, and its provision reduces the cost of production. The first paper assumes that the local public good decreases the fixed cost of producing a variety. The second assumes that it reduces the marginal cost of production. The impact of the two types of local public goods differs. A local public good that reduces the marginal cost of production also reduces the selling price of the manufacturing varieties, since the selling price is a mark-up on the marginal cost of production. This policy generates inter-regional spillovers, as it reduces the aggregate price index in both regions. It is also a correcting device for the monopoly pricing of the manufactured firms: marginal cost subsidies drive prices towards the marginal cost of production. Nevertheless, as expected, the supply of a local public good also attracts firms. This type of public good has the advantage of alleviating possible redistributive effects, as it generates a significant amount of positive spillovers.

A local public good that reduces the fixed cost does not have any direct impact on the pricing of firms. Its main effect is on the equilibrium number of varieties that are supplied. Supplying more public good in one region increases the number of varieties that are supplied locally and reduces the number of varieties supplied in the other region. The effect on the total number of varieties is ambiguous and depends on the relative fixed cost of production of firms. Again, the provision of this type of public good affects the inter-regional distribution of production.

1.3.2 The choice of equilibrium

Economic geography models help to explain why apparently similar regions may experience radically different economic fortunes. Inherently, there is a multiplicity of equilibria. In this section, we investigate what is the most likely equilibrium outcome and whether public policy can select the best one.

We have already eliminated the unstable equilibria, but the economies described in this literature are left with several stable equilibria. When agglomeration takes place, all regions are potentially the site of the core. Moreover, for some ranges of parameters, we may have two different types of equilibria: two core-periphery equilibria and one dispersed equilibrium. In that case, the selection of equilibrium also matters in terms of equity. If a policy induces a switch from one equilibrium to another, the switch should persist over time.

The literature adopts different points of view concerning the selection of equilibria. On the one side, there are the proponents of the historical determination of equilibria. Historical events set conditions that drive the economy to one or another core-periphery equilibrium. Krugman (1992) discusses and explains different historical events that may have induced such a selection. For him, the agglomeration of the carpet industry in the US in Dalton (Georgia) dates back to a famous bedspread that was given as a wedding gift in 1895 by a teenage girl, Catherine Evans Whitener, to her brother. She revived a technique that had been developed locally and because of Marshallian externalities it induced the agglomeration of the industry there (and not in another place). He also explains why an important part of the industrial activity developed in the rust belt, due to the arrival of new immigrants from the Northern US and the local availability of natural resources. More relevant to our concern, he explains the sudden development of California in the mid-19th century as a long-term effect of the 1848 gold rush that suddenly displaced workers and demand from the North to the South.

On the other side are the approaches that focus on the role of expectations. The equilibrium that is going to prevail may simply depend on the expectations firms and workers form about which region will be the most developed in the long-run. If adjustment costs are nil, there would be no reason to believe that history matters. Workers/firms would adjust instantaneously to changes in expectations. If adjustment costs are positive, the decision for workers to migrate or for firms to relocate would be an investment. As Krugman (1991a) puts it, whether expectations or history matters then depends on the particular dynamics involved in the model. Recent economic geography models are rather silent about the specific dynamics involved.

Borck et al. (2009) analyse the public policy implications of multiple equilibria. The authors propose a footloose capital model with two regions differing in market size. There are both internal scale economies (i.e., firms produce under increasing returns) and external scale economies (e.g., knowledge spillovers). In such a framework, in the absence of external returns to scale, the equilibrium entails the large region hosting the core (with either full or partial agglomeration). However, in the presence of knowledge spillovers, and if they are sufficiently strong, it may well be that the small region becomes the core (although the core in the large region is also an equilibrium in that case). This is inefficient inasmuch as there are more consumers in the large region, who must pay higher, periphery prices due to positive trade costs than in the small region, where consumers pay lower, core prices. Borck et al. (2009) ask whether capital subsidies can lead the economy out of the inefficient lock-in. They show that introducing a capital subsidy in the big region may destroy the inefficient equilibrium.

Kline (2010) develops a similar argument in a knowledge spillovers model. He shows that in this context, the intervention of the government has the potential to reverse the agglomeration process. The provision of a capital subsidy can have substantial and long-term effects. For the policy to work, it must change the agents' expectations. The sole announce of the subsidy may act as a coordination mechanism for the agents of the economy to switch from one equilibrium to another.

Chapter 2

The gains from spatial concentration (empirics)

Economists have long endeavoured to evaluate the magnitude of the gains from spatial concentration. Typically, the relationship between local productivity and the size of the local economy, or the determinants of spatial concentration have benefited from numerous and insightful contributions (see the survey by Rosenthal and Strange (2004)). Nevertheless, this area has recently undergone a kind of revolution, due to two major events.

First, whereas previous studies had considered very large spatial units (states in the US, countries or NUTS1 regions in Europe), new data sets became available for much finer geographical classifications. The very design of the classifications was improved, changing from purely administrative definitions to economic ones that better match the reality of the location mechanisms at work. For instance, a number of European countries have defined "local labour market areas" (LLMAs): consistent areas where people live and work, where daily transboundary commuting is limited. This truly corresponds to the scale at which a number of agglomeration mechanisms occur, in particular those related to better matching on local labour markets, the diffusion of ideas and innovation and the use of non-tradable goods and services. Importantly, it does not prevent the researchers from simultaneously considering effects taking place at larger geographical scales, related to regional or international trade, that are highlighted by recent economic geography models.

For example, 341 LLMAs were defined in continental France, completely covering it, whereas there are only 5 NUTS1 regions and 21 NUTS2 regions. Italy defined 686 LLMAs for only 21 Nuts2 regions. What can be captured econometrically with such an improved geographical focus is clearly more relevant for both economic analysis and economic policy.

Recent years have witnessed a further evolution with the use of individual data (on workers or firms) and the design of empirical specifications that encompass both aggregate spatial effects (typically evaluated at the LLMA level) and individual effects, whereby the individual heterogeneity that exists within regions can be removed from the estimates of LLMA effects.

The second development that has transformed empirical research in economic geography is the improvement of econometric methods themselves, particularly the more systematic use of panel and instrumental variable estimation techniques. They enable researchers to take into account the individual heterogeneity that cannot be estimated by the variables available in data sets, and to deal with missing

variables and reverse causality. The latter are particularly crucial in economic geography because of the snow-ball effects mentioned above. As we shall describe in more detail below, endogenous location choices basically make all variables studied endogenous, making it impossible to use standard econometric techniques.

A final trend consists in a systematic effort to improve the relation of econometric specifications to theory, in order firstly to make more accurate interpretations and secondly to better identify possible missing variables and reverse causality issues. This has led to a fairly sharp divide in the nature of papers produced in empirical economic geography. The first strand of literature remains fundamentally "reduced-form". The specification considered includes all possibly relevant variables, although it is impossible to fully write and design a theoretical model that would lead to this specification. These studies can be the most useful for policy-makers, since all variables of interest can be included in the specification.Nevertheless, this usefulness has to be qualified because of the interpretation problems it raises. Some of the variables may be interdependent, and this is not taken into account in the specification. Consequently, some of the effects cannot be completely estimated. Given the number of variables introduced into the regression, endogeneity biases may increase, but at the same time they are difficult to identify, since no theoretical model supporting the specification is available.

The second strand of literature follows the opposite route. It starts by designing a theoretical model, and then derives the specification to be estimated from that model. Consequently, one knows precisely which effects are studied in the empirical model and which are not, and which are the exogenous and endogenous variables. The drawback is that in general, far fewer effects can be considered simultaneously, so that these studies only provide partial answers for policy-makers. In our survey, we will draw on both bodies of literature. In this chapter, we start by reviewing the various possible empirical strategies, firstly reduced-form and then structural, before moving on to discuss the results obtained. Importantly, in the final section we list all the possible regional policy implications that can be drawn from this material, even if it was not originally designed for such a purpose. The next chapter deals with the studies specifically designed to evaluate the role of regional policy.

2.1 Empirical strategies

2.1.1 Wage and productivity equations

Reduced-form approaches, theoretical framework

In economic geography, even more than in other fields, it is essential to define the theoretical background of reduced-form empirical approaches as clearly as possible, because of the interdependency between local economic outcomes and agents' location choices. It has been shown by Combes, Duranton and Gobillon (2008) that the literature reviewed in this section fits into a very simple framework. This is not a complete economic geography model; nevertheless, it sheds light on the relationship between productivity and the industrial characteristics of the region of interest that is being estimated.

Taking the example of the impact of density on productivity, and following Ciccone and Hall (1996), which is a milestone in the new departure of this literature, researchers typically regress the average hourly wage in the region-industry on total employment density, the error term representing what is left unexplained by density. The specification is usually estimated in logarithms, allowing the researchers

to estimate the productivity elasticity with respect to density, which is assumed to be constant. This approach can be used, for instance, to evaluate the part of the productivity gap between more- and less-developed regions that is due to their size differences. It also allows to estimate the potential impact of increasing or decreasing the size of a region. However, one must be very cautious when it comes to drawing policy implications from such figures, as the Combes, Duranton and Gobillon (2008) model shows.

This model assumes that in any region and industry, firms operate under constant marginal costs using labour and another input summarising the role of any input (primary or intermediate) other than labour. The overall level of technology and labour efficiency are firm-specific. A firm's profit is the sum of the profit it makes on all the markets on which it operates, bearing in mind that it incurs trade costs for distant markets. The profit function can be written as the difference between the firm's total income and the labour and other input costs. Total income is the product of the firm's production and of its income per unit produced, which can be called, a little bit abusively, the "price" of the good (for the firm). It is the average, over all the markets on which the firm operates, of the money the firm makes on each unit sold, net of trade and intermediate consumption costs. Typically, it is higher when the firm is close to large markets because the trade costs incurred on both the good sold and intermediate consumption are lower. This is part of the agglomeration effects we described in Chapter 1, to which economic geography gives micro-foundations. However, as long as one estimates reduced-form specifications, these mechanisms do not need to be detailed. Conversely, equilibrium prices are lower when competition is strong, which is more often the case in large regions, and this now acts as a dispersion force, as again highlighted in economic geography models.

Even assuming perfect competition on both goods and inputs markets, which is a huge simplification by comparison with economic geography models, Combes, Duranton and Gobillon (2008) show that writing the first-order condition of the optimal use of labour by firms, which equalises each worker's wage with their productivity, supports a wage specification to estimate agglomeration effects. The nominal wage a worker gets is first directly proportional to his own skills and then depends positively on both the firm's technology level and the price of its good, and negatively on the cost of inputs other than labour. Alternatively, one can work on total factor productivity instead of wages as the dependent variable, since it can be shown to depend on the same variables.

These equations can be used to summarise most economic geography effects, allowing us both to justify the inclusion of most RHS (independent) variables considered by this literature and to interpret their effect. Typically, productivity (and therefore wages or total factor productivity) will be high in a region that benefits from technology or labour efficiency pure externalities, whether they consist in technological spillovers or improved matching on local labour markets. But wage and total factor productivity are also high if market access is good relative to the degree of competition, resulting in high prices net of trade and intermediate consumption costs, which is strengthened when competition is imperfect. Finally, firms may also benefit from Marshallian externalities working through local input markets. If the access to inputs other than labour is good, resulting in low costs, then wages and total factor productivity are also high. These causes and effects can also be read in the opposite direction. Productivity will be low if pure congestion effects dominate in the region, worsening the technological level or the labour efficiency, for instance when local transport networks are saturated. The same result arises if market access is bad, either because competition is strong or because markets are distant, or if access to inputs other than labour is bad, land being an extreme example. Due to the very rigid local supply, land price is one of the

input prices that increases the most with the size of a region.

By comparing this model with the econometric specification between productivity and density that has been estimated, one can obtain the correct interpretation of elasticity. Assuming that the average marginal productivity of all firms located in the same region-industry is proportional to density means, according to the model, that density either has an impact on productivity, through the overall technological level of the firms, labour efficiency, or the price of the good (a positive impact, i.e. density increases these variables which in turn increase productivity) or it has an impact on the costs of inputs other than labour (a negative impact, i.e. density increases these costs which in turn reduce productivity). Importantly, the model underlines that such estimations do not identify the channel of agglomeration economies (technology, labour efficiency, goods or input prices) separately. Only the overall impact of these effects is evaluated. Moreover, only the net effect of density is identified, since positive effects may be partly offset by negative effects, and vice versa. Nevertheless, as we explain below, estimation of this overall net impact of density on productivity is crucial from the policy point of view, and already provides many clear conclusions.

Local productivity is usually explained at the industrial level. The dependent variable is the regional productivity in a given sector, and the observations for all sectors are pooled together for the estimation. Two types of variables are included in the regressions: variables measuring characteristic specific to region and the one measuring characteristics specific to the industry. The first type of variable includes the total employment density, local industrial diversity or the market potential. The type type of variable include the degree of specialisation of the industry or the average size of firms operating in the industry.

Region specific variables

Total employment density is measured at the regional level and reflects what are called "inter-industry" (or "urbanisation") externalities, which are the effects of the overall characteristics of the region, which do not depend on the industry. Other proxies for inter-industry externalities that are usually included in such regressions are the overall size (land-area) of the local economy and local industrial diversity, which measures the role of the distribution of activity across local industries, typically using a Herfindhal index on local industrial shares. The former captures the fact that a "large" region can be large either in terms of spatial extension, hence the land area, or in terms of thickness, which is captured by density. The two can affect productivity simultaneously, but for the policy-maker increasing land area or density are two somewhat different strategies. It is therefore important to evaluate both their roles. Some authors introduce directly total employment instead of these two. Interpretation is made more difficult by the fact that total employment captures both the thickness and the area effects of local size.

Diversity can encompass two very different kinds of agglomeration channels. Long ago, the geographer Jacobs (1969) argued that cross-fertilisation of ideas between industries can occur and be a driving force for new innovation. In this case, regions where industrial diversity is high would benefit from stronger growth. Alternatively, and as argued in section 1, in economic geography models based on monopolistic competition with Constant Elasticity of Substitution utility functions or inputs, industrial diversity is also the main driver of agglomeration, through completely different mechanisms. This is the second type of effect that the diversity index effect would capture in these regressions.

Market potential, as shown in Figure 4, has been added to these specifications in recent years. The

aim is to capture the fact that when goods are tradable, it is not only the size of the region's own local economy that matters, but also access to more distant markets. This captures the same group of effects as density or land area but emanating from neighbouring regions. In other words, because of the intense interactions that take place between regions, there is no reason for agglomeration effects not to spill over spatial boundaries, at least partly. The external - not total as mapped in Figure 4 - market potential is used. Own-region density is not included because its effect is entered into the specification separately. Again, this enlarges the spectrum of possible policy strategies evaluated, since the effects both of own-region size and of access to other regions are estimated. Some authors go even further in the attempt to identify the spatial decay of agglomeration economies, introducing the regional characteristics separately at various distances from the region. They start, for instance, with the own-region density, then the density of regions less than 50 kilometres away, then the density of regions between 50 and 100 kilometres away, and so on.

Industry-region specific variables

The second type of variables includes in the regression measure the industry's local characteristics that may influence productivity. This concerns externalities operating within the industry, known as "intra-industry" (or "localisation") externalities. The most standard intra-industry externalities proxy is the degree of specialisation of the region in the industry, measured by the share of the industry in the local economy. Glaeser et al. (1992) popularised another terminology for these externalities, using the term "Marshall-Arrow-Romer (MAR) effects" (in reference to those who first evoked such an intuition) for the role of specialisation and "Jacobs" externalities for the positive role of diversity. The local average size of firms, which can capture the magnitude of intra-firm economies of scale, is also sometimes considered in the right-hand-side of the specification. The last explanatory variable is the share of professionals in the region-industry. This evaluates the possible role that such workers can play as a conduit of technological spillovers. A relatively complete review of all the variables introduced into the specification can be found in Rosenthal and Strange (2004).

There is another reason to control in the specification for the skills composition of the labour force. In local productivity, it is important to isolate those effects that really result from local externalities from more direct differences in labour efficiency across occupations - differences that would be present irrespective of the location. If the uneven spatial distribution of occupations is not properly controlled for, then some regions could appear more productive even when there is no local externality, simply because they host more skilled people. While it is interesting to understand why the distribution of skills is uneven, a question to which we shall return below, the literature proposes strategies to quantify the presence of local externalities separately from this composition effect of local labour markets. At the very least, as Hellerstein et al. (1999) suggest, the share of the various occupations or education levels in the local labour force must be introduced into the specification. When individual data is available, one can consider the own-skill level of each worker separately from the aggregate externality effect on productivity of each type of skill in the region. Otherwise, only the combined result of these two effects is identified.

Finally, this literature acknowledges the fact that local endowments, in a broad sense, do differ across regions, and that this affects productivity, again possibly independently from the presence of agglomeration economies. Typically, one would like to control for private and public capital local endowments

(research centres, universities, transport infrastructure), which can be under the control of policy-makers, physical geography (access to coast, river, central location, etc.) and for the quality of local institutions or the level of available technology. The last two are more relevant in the context of developing countries (including China) and for European regions belonging to countries with differing institutional designs (possibly due to historical reasons), and probably less relevant within well-integrated European countries. Unfortunately, as we will see below, the role of each of these components is rarely evaluated separately, but their overall effect can be controlled for to obtain the impact of density net of these endowment effects. In other words, what is estimated is not the fact that denser areas host more public goods, but the impact of density for a given provision of local public goods.

Note, as both Combes, Duranton and Gobillon (2008) and Moretti (2004a) emphasise, that when assessing productivity by wage rather than total factor productivity, the dependent variable is the nominal wage, not the real wage. There is no reason to deduce the cost of living (and in particular land and housing costs) from nominal wage nor to control for it in the right-hand-side of the specification. What is estimated is a labour productivity equation, dealing with nominal return, and not a migration equation that would deal with real income differences between regions.

Structural approaches of wage-equations

We have underlined that reduced-form specifications of wage or productivity neglect the richness of economic geography models. For instance, trade costs are only crudely incorporated, since market potential is the only variable that takes into account the trading possibilities between regions. Economic geography models make an important contribution by highlighting the fact that in general equilibrium, trade costs affect most endogenous variables: not only quantity and prices but also factor returns and endowments. Consequently, in empirical exercises, almost all explanatory variables should depend on them, bringing into play numerous direct and indirect effects that cannot be disentangled without a more precise theoretical background to underpin the empirical specification. For this reason, an empirical literature has been developed to take theory even more seriously and to use it to derive wage specifications directly obtained from theoretical models. This allows the researchers to consider more channels through which economic policy variables affect regional disparities.

Typical models used to sustain structural estimations are borrowed from the Dixit and Stiglitz economic geography approaches, more precisely the Krugman (1991b) and Krugman and Venables (1995) settings that were mentioned in Chapter 1. The first step consists in extending these models to frameworks encompassing a large number of regions and industries. Manufacturing goods are differentiated into varieties and consumers have CES preferences over these varieties. In addition to labour as an input, the production function may include a CES composite input made of the manufacturing varieties. Exports are subject to trade costs. Monopolistic price competition with free entry takes place between firms.

Two main structural empirical strategies have been envisioned. The first, proposed by Hanson (2005), is probably more relevant for small, within-country, geographical scales. Agglomeration mechanisms are based on labour migration, people being perfectly mobile across locations. The Redding and Venables (2004) approach deals with situations where labour spatial mobility is low (it assumes no inter-regional mobility) and emphasises the role of intermediate inputs. In both cases, it is shown that wages, and therefore local labour productivity, are functions of market potential. However, this market potential,

while close in its spirit to Harris' market potential, encompasses more sophisticated effects. Generally speaking, market potential corresponds to the intuition that firms located in or near a large region can access a larger market than those located in a small region. This is because on distant markets, trade costs make firms less efficient than local firms. Firms have higher market shares on nearby markets and therefore sell more when these are large. In other words, the size of the firm's market corresponds to a spatially discounted sum of all market sizes (in terms of population or employment), which is what Harris' market potential emphasises. Now, in a fully-fledged model under imperfect competition, it is easy to see that the magnitude of the market share loss on distant markets due to trade costs depends on the degree of asymmetry between markets, in terms of both the number of firms and the demand for variety, which are endogenous in the model. In equilibrium, the firm's total sales depend not only on the spatially discounted market size but also on price effects reflecting the extent of competition on each market. Typically, in such settings, productivity is shown to be a function of what researchers call the real market potential, a Harris market potential corrected by price effects. It depends on price indexes in all locations, which are CES functions of variety prices. The main problem for the econometrician consists in dealing with these price indexes. On top of being endogenous, they are highly non-linear in unknown parameters, which make them typically unavailable in data sets.

Hanson (2005) tackles this issue by considering the Helpman (1998) extension of the Krugman (1991b) model. He gives a role to local housing markets, which allows him to escape the non-realistic equilibrium where industry completely disappears from some regions. Migration decisions depend not only on nominal income and variety price indexes but also on housing prices. Using the fact that under perfect mobility, indirect utilities equalise across locations, and considering all other markets clearing conditions, the price indexes can be replaced in the real market potential variable by observed variables, namely local wages, total income and housing stocks. Importantly, this structural model therefore shows that price effects in a fully-fledged economic geography model modify the specification of the market potential variable, making it necessary to include variables in the market potential function other than those considered by Harris (1954). Note however, in order to use those observable variables as proxy for market potential one have to believe in the model and therefore implicitly assume away all impact that, among others, local public good or amenities could have on the location decisions.

The approach is structural because the specification estimated is directly derived from the theoretical model and because estimated parameters can be related to the theoretical model's parameters (the elasticity of substitution between varieties, the share of manufacturing in consumption, and trade costs) for which estimates are therefore also obtained. The Redding and Venables (2004) approach is also structural but it emphasises different agglomeration and dispersion forces. There is no longer any role for the housing market, but a new effect working through intermediate inputs emerges, which also takes the form of a real market potential variable. Moreover, this is the estimation of a trade equation in a first step, which is not very data-demanding, that allows the authors to obtain predictions of the real market potential variables on which wages are regressed in a second step.

2.1.2 Employment growth, firm creation and local dynamics

Apart from Ciccone and Hall (1996), two other studies were also fundamental in launching the empirical evaluation of agglomeration economies in the 1990s. These are Glaeser et al. (1992) and Henderson et al.

(1995). The main difference with Ciccone and Hall (1996) concerns the dependent variable explained. Instead of seeking to explain why some regions are more productive than others, Glaeser et al. (1992) and Henderson et al. (1995) evaluate the local determinants of local employment growth. This is also very important to policy-makers, especially in situations where regional unemployment disparities are large, as they are in Europe.

Using the same theoretical model as for productivity equations, we can again find some theoretical foundations for the specification estimated. We will use the framework of economic geography to analyse and interpret the different effects measured in those contributions. It is fair to note that Glaeser et al. (1992) had more the endogenous growth models in mind.

The first-order condition relating wage to labour productivity that was used before can be inverted to get a specification with labour on the left-hand-side and wage on the right, together with the same variables as before, i.e. the good's price, the technology level, the labour skill level and the cost of inputs other than labour. Therefore, beyond a number of concerns identified by Combes (2000) about the way the agglomeration variables are defined, making the same assumptions on how inter- and intra-industry externalities affect this group of variables, it is possible to identify the same effects (of density, diversity, market-potential, specialisation, etc), but now on local growth.

Importantly, the model proposed by Viladecans-Marsal (2004) shows that it is necessary to control not only for wage but also for production. This may raise serious econometric issues, as we will see below. Not controlling for wage and production as many authors do is possible, as illustrated by the model proposed by Combes et al. (2004), but it changes the interpretation of the agglomeration effects estimated. For instance, it is shown that in this case a positive agglomeration effect on productivity (of density for instance) only has a positive effect on employment under certain conditions on the demand and labour-supply elasticities. This is intuitive: if demand is not very elastic, a positive local externality that increases productivity but not production will save on labour, causing local employment to decline. Hence, when a negative effect of specialisation on employment is estimated, for instance, the policy-maker does not know whether specialisation has a negative effect on productivity, and therefore a negative effect on employment because productivity and employment are positively correlated, or if specialisation actually has a positive effect on productivity, but productivity then has a negative effect on employment. This may limit the usefulness of the estimation of the effect of specialisation on local growth.

As another extension, Combes et al. (2004) break down the growth of local employment into two terms, the growth of employment per firm and the growth of the local number of firms. This allows them to evaluate the role of regional characteristics on employment in existing firms and on firm creation separately. In other words, density or specialisation can have different effects on the intensive and extensive margins of employment. For policy-makers, this indicates whether the same tool has the same or opposite effects on internal growth (enhancing local firms' growth) and on external growth (new firms are created or less firms are destroyed in the region), or whether it is more efficient to target one or the other. Some authors, such as Rosenthal and Strange (2003) on US data, concentrate on the determinants of firm creation.

One strand of literature somewhat combines the two types of studies (productivity and employment growth). It uses data on production, as in section 2.1.1, but estimates the determinants of production growth, in the spirit of Glaeser et al. (1992) and Henderson et al. (1995). It has the advantage of reducing some of the endogeneity issues and of clarifying interpretations we detail below, for instance due to the

unclear effect of productivity gains on employment. In this case - and it was also proposed for employment growth by Henderson (1997) and Combes et al. (2004) -, the authors, including Henderson (2003) and Cingano and Schivardi (2004), estimate full dynamic models. Production or employment in the region-industry at a given date is explained by its past values and both the current and past values of inter- and intra-industry externalities. This has both econometric and policy advantages. From the econometric point of view, the dynamic panel data methods (a short-cut used here to refer to the General Methods of Moments applied to panel data, following the literature initiated by Arellano and Bond (1991)), allow the researchers to control for possible endogeneity concerns with minimal data requirements. We will examine this in more detail below. Moreover, and interestingly for the policy-maker, an assessment is obtained of how long agglomeration effects last. It can tell us, for instance, for how many years a local increase in density will raise productivity.

2.1.3 Location choices and logit models

Instead of explaining local productivity or local growth as a function of local economic characteristics, other authors have sought to evaluate the impact of these characteristics directly on the location of firms. Carlton (1983) proposed to use, for this purpose, the discrete choice modelling strategy developed by McFadden (1974). Economic geography predicts how firms distribute themselves across space according to the relative accessibility of the regions under consideration, after controlling for the differences in regional costs. Any of the local characteristic variables we mentioned above can be considered to assess both groups of variables. As for productivity equations, both reduced-form equations, as in Carlton (1983), and structural equations, as in Head and Mayer (2004), can be envisioned, with the same advantages and drawbacks.

This type of research has primarily been applied to multinational firms, because the determinants underlying their location decisions are more readily discernible than those for domestic (and therefore less "footloose") firms. In particular, multinational firms location choices occur over the course of a relatively short time period, and they are free from the historical contingencies to which national firms are often subjected.

Here again, sketching, even briefly, a theoretical model that can sustain the specification estimated is very useful as a means to assess possible econometric issues and to clarify interpretations. A firm wants to establish a subsidiary in the region that can provide the highest profits. If regional profits were ranked in the same way by all firms, and if they all chose their location at the same time, a single region would end up attracting all firms or, at the other extreme, all regions would attract exactly the same number of firms. This is what economic geography models predict when they do not consider any random component, but it is clearly not what is observed in reality. It is actually McFadden (1980)'s great contribution to propose a way to escape this paradox for any discrete choice. He suggests that discrete choices reflect two components of a firm's or individual's preferences: a deterministic one, shared by all agents, and a random one, specific to each agent. Applied to location choices, this means that the profit a firm gets in a given region is made up of the profit that any firm would get in this region plus a random term specific to the firm. Each firm draws its specific component for a given region independently from other firms and independently from the components it draws for the other regions. As a result, the ranking of these random components across locations differs for each firm, and therefore the ranking of

the profit they get in each region also differs. They do not end up choosing the same region. Clearly, the larger the variance of the random component relative to the mean of the deterministic component, the more even the spatial distribution of the firms, and the smaller the variance, the more concentrated the spatial distribution.

Many rationales can be found for the presence of the random term. Ffor example, depending on their history, or the presence of other firms belonging to the same group, some firms are more efficient than others in a given region, hence the specific profit they can make there. The second contribution of McFadden (1974) was to show that under some assumptions on the random term distribution, the effect of variables that determinate the discrete choice can be easily estimated using logit models. Since it is quite easy to show that the same variables that determine regional productivity affect profits in the same way, researchers estimate multinomial models where the probability of locating in a region, empirically assessed by the share of firms located there, is a function of local characteristics similar to those used in productivity equations.

Conditional Logit models assume that the random terms are independently and identically distributed across regions. This is a strong assumption because some may be located closer to each other, or belong to the same country within a larger area, like Europe. A first solution to escape this extreme assumption consists in including control variables at the country level for some characteristics - institutions, taxes, technological development for instance - that are common to groups of regions (belonging to the same country, for example). Alternatively, a specific strategy, called the Nested Logit model, assumes that location choices are made in two stages. Firstly, firms choose a country in which to locate, and then, conditional on this choice, they choose the region. Two random components are considered now, one specific to the region, one specific to the country, and they are assumed to be independent. However, within a country, random terms are correlated to an extent that is estimated simultaneously with the model's other parameters. Hence, by putting more structure on the way location choices are made, this strategy allows for a more sophisticated structure of the random component. The location choice determinants at different spatial scales are evaluated separately, once the geographic decomposition has been chosen (country then region, for instance, but it could be Western or Eastern Europe then region, and so on). Interestingly, a statistical test is provided to assess which model, nested or not, is preferable. The reader can find in Train (2003) a detailed presentation of the techniques for estimating this class of models.

2.1.4 R&D, innovations and technological clusters

If population or employment is on the whole concentrated, another well-established fact regards the even stronger concentration of some specific activities. R&D activities and their outcome, innovations, have long been documented as presenting strong spatial concentration patterns, as testified by a number of case studies described by Porter (1990) or Saxenian (1994), for instance. This spatial concentration is confirmed using harder evidence by Feldman (1994) or Audretsch and Feldman (1996) on US states. This concentration is often seen as a proof in itself of the existence of spatial technological spillovers. According to this idea, distance limits the spatial diffusion of knowledge, which in turn encourages firms, who imitate from each other, to locate in the same areas.

This relates to one of the gains from agglomeration that Marshall (1890) had already mentioned.

Because they exchange more ideas and knowledge than when located far apart, firms and workers located in the same areas innovate more. A specific empirical literature attempts to directly test for the presence of this precise type of agglomeration effect. By comparison with the impact of density on productivity or of past foreign firm presence on FDI location choices, which can testify to the presence of agglomeration effects but cannot identify their channel, this is really one of the mechanisms behind agglomeration effects that authors seek to bring to light here. Typically, the authors estimate what is called from Pakes and Griliches (1984) a knowledge production function, typically the number of patents the firm or the region produces as a function of the firms or regional inputs, especially R&D, and that can encompass the presence of some local externalities emanating from R&D performed in local/nearby universities or local/nearby firms. The impact of the local economic structure (density, specialisation, diversity) on the strength of such effects has also been studied. Alternatively, researchers have also tried to show the role of distance on the diffusion of knowledge by comparing the location of the patenting activity and the location of patent citations.

2.1.5 Econometric issues raised by endogenous location choices and spatial selection

Clarifying the theoretical framework as researchers have endeavoured to do in most recent studies is important for an accurate interpretation of the results. Such clarification is also very useful for assessing whether or not econometric estimates suffer from endogeneity bias. Typically, the simplest estimation strategy, based on Ordinary Least Squares (OLS), assumes that no explanatory variable is correlated with the residual of the specifications. No variable that influences the local economic outcome studied and is not included in the specification must be correlated with the variables that are actually introduced into the specification. Given that local economic characteristics are often inter-dependent, if only through the location choices of firms and workers, this is not such an easy assumption to satisfy. There are two possible sources of correlation between explanatory variables and the residual, and therefore of endogeneity: missing variables and reverse causality. We start by illustrating them in the case of the effect of density on productivity, but they concern almost all the variables that this literature considers, both dependent and explanatory.

Missing variables and reverse causality

A missing variable is a variable that does actually affect the phenomenon studied - local productivity, local growth, firms location choices, innovation - but is not included in the specification. Possible missing variables are numerous. In the case of the link between productivity and density, for example, some industries are more productive than others at the national level and at the same time over-represented in the densest regions. This is also the case for certain local endowments, such as local public goods or geography, that increase productivity. If such variables are correlated with density and omitted from the regression, they are captured by the error term, which is thus correlated with one of the explanatory variables. The effect of density, interpreted as the presence of local externalities, is estimated with a bias because it captures effects that do not directly reflect agglomeration economies. In this case, dense areas are more productive not because density favours local interactions that make workers and firms more productive, but because dense areas benefit from the presence of the most efficient industries or of local public goods that increase productivity. As another example, areas that are not dense may

suffer from geographical elements (mountains for instance) that decrease productivity. This also creates a positive correlation between density and the residual (they are both low in these areas) and also leads to a positive OLS estimate of the impact of density on productivity that is not related to agglomeration effects. Correctly interpreting the impact of density is crucial for regional policy, because if it is local public goods, and not density itself, that are generating productivity gains, then locating more public goods in the periphery would increase productivity there. If the reverse is true, density but not local public goods increase productivity, locating public goods in peripheral regions would not increase productivity there. Therefore, when using OLS estimates, one must be sure that all the variables that have an impact on the dependent variable are introduced into the specification, otherwise OLS should not be used.

Note that the estimation bias due to missing variables can work both ways. Density economies are under-estimated if one omits variables in the regressions that are negatively correlated with density but enhance productivity, or, which is more frequent, variables that are positively correlated with density but have a negative impact on productivity. For the latter case, Roback (1982) proposes an interesting model, showing that only careful analysis allows the researcher to assess all the variables that need to be controlled for in the specification. Roback (1982) first states that some consumption amenities, such as cultural goods, leisure facilities or restaurants, are both over-represented in cities, and therefore positively correlated with density, and attractive to people. Households, who move to cities for this reason, consume housing, thereby increasing land and housing prices. These higher land prices give firms an incentive to substitute labour for land in the production process, which in turn reduces the marginal productivity of labour. Therefore local consumption amenities, although not a direct determinant of productivity, have to be included in wage/ productivity equations, otherwise the impact of density, which would include their negative impact, would be under-estimated relative to the true value corresponding to agglomeration effects. Only a model like the one developed by Roback (1982) allows us to realise this.

Another sort of missing variable problem relates to the fact that some areas may select agents that are not identical to those selected in other regions, as discussed in Moretti (2004a) and emphasised in Combes, Duranton and Gobillon (2008). For instance, imagine that high-skilled workers have stronger preferences for city amenities than low-skilled workers. Or that high-skilled workers benefit more from agglomeration economies than low-skilled workers and, anticipating this, they are more likely to locate to dense areas. In that case, part of the higher productivity observed in denser areas is only due to the over-representation of high-skilled workers in the area, since they are more productive wherever they locate to. If one does not control for this selection problem, for instance by at least introducing variables that reflect the skill level of local employees, one tends to over-estimate the impact of density, which also captures this labour composition effect.

The second source of endogeneity, reverse causality, arises when people and/or firms choose their location according to the returns they get there, which are in general directly linked to the dependent variable studied. In the case of productivity, for instance, productivity shocks unobserved by the econometrician but observed by workers and/or firms become correlated to density, which now depends directly on these shocks due to endogenous location choices. OLS estimates are again biased. For instance, imagine that certain local government authorities choose more efficient policies, or are run by cleverer politicians. This is not recorded in the data sets used by researchers, but agents who are looking for a new location may learn about the presence of this more efficient government and therefore of the higher gains they will make in this region through discussions between entrepreneurs or workers or articles in newspapers. This

positive productivity shock in the region gives them an incentive to move there, which in turns increases density there, hence the positive correlation between the two. In this case, it is not density that increases productivity, but productivity that increases density. The causality is reversed; hence the term "reverse causality" used to designate such a situation. Unfortunately, OLS estimates of the effect of density on productivity are positive in this case, even in the absence of positive agglomeration effects, because density captures the positive impact of the government. As another example, some regions may be hit by bad climate events (with extremes like hurricane Katrina in New Orleans). Again, the negative impact on productivity may be taken into account in the location choices of firms and households, but not recorded in the data set the econometrician uses. Agents moving out of or not moving into these areas cause the density to fall there, where productivity is also lower. Again, this creates a positive correlation between the two variables, and OLS estimates of the impact of density on productivity will be over-valued.

Fixed-effects and instrumentation solutions

Fortunately, both sources of bias can be circumvented. One strategy consists in introducing all available control variables, as long as they are consistent with the theoretical framework. For instance, one can imagine data sets on the location of public goods that affect firms' productivity, such as airports, high-speed train stations or universities, or of consumption amenities such as hospitals, cultural amenities, restaurants and so on. The average education- or skill-levels of the regions are also sometimes recorded in data sets. Unfortunately, the list of possible controls is so long that one may doubt whether econometricians would be able to gather all the data needed. If panel data are available, fixed-effect strategies can be adopted, for instance to control for aggregate effects such as nation-wide, industry-specific effects. A fixed effect is a dummy variable that take value one for all observations that correspond to the same sector, or more generally to a group of workers or firms. The effect of this variable captures the impact of any variable specific to this group. The advantage is that it is no longer necessary to gather data for all such variables. The drawback is that the effect of each component of the fixed effect is not identified separately. In other words, fixed effects control for a group of effects but are in general difficult to interpret due to the number of effects they cover. They are therefore difficult to use to draw policy implications, but they allow for a correct estimation of the effect of the other variables simultaneously included in the specification, provided that they vary in another dimension - time or space for instance - than the one controlled by the fixed effect. Nevertheless, as demonstrated in Combes et al. (2011), using fixed effects may slightly change the interpretation of the results, as we will illustrate below.

One of the significant advances made possible by individual panel data sets is the possibility of taking individual fixed effects into account. As described by Glaeser and Maré (2001), Moretti (2004b) and Combes, Duranton and Gobillon (2008), when the same workers are present for many years in the data set, one can identify separately the role of their overall time-invariant characteristics (with an individual fixed effect, a dummy variable taking value 1 for all the dates at which the individual is present in the data set) and the role of the density of the region where they are located, which changes over time. The estimated impact of density is then net of the effect of any individual skill, be it the education level of the worker, the education and overall background of his parents or grand-parents, the role of the city, state or region where they were born and where they lived afterwards, the family structure before the date of observation, etc. Among other things, this strategy corrects for the fact that some regions may select

certain workers who are different from those located in other regions.

The second strategy for correcting the OLS endogeneity bias relies on what is called instrumentation. The aim is to find "instrumental variables" (IV), or "instruments", correlated with the possibly endogenous explanatory variables - typically density -, but not with the shocks that affect the dependent variable and therefore enter the error term. One then replaces in the regression the possibly endogenous explanatory variable by its predictor obtained from its regression on the instruments. This predictor is not correlated with the error term because it is proportional to the instruments that are by definition not correlated. The OLS estimate of the predictor's impact on the dependent variable can be shown to be unbiased and identical to the true impact of the explanatory variable on the dependent variable.

The difficulty is to find the instruments, which have to be correlated with the instrumented variable and at the same time uncorrelated with the error term, while the instrumented variable is correlated with this error term, creating a tension between the two objectives. Ciccone and Hall (1996) propose to use historical variables to instrument density, for instance local density or population many decades before the years for which the model is estimated. Due to the inertia of the urbanisation process, historical and current regional densities are still fairly well correlated, since present-day large cities were already often the largest cities centuries ago. If this is correlated with physical geography and physical geography still affects productivity and is not included in the regression, then these historical variables are bad instruments, since they are correlated with current productivity shocks. But if shocks do not correspond to such permanent effects, for instance correspond to recent public goods or to technological shocks that took place after the date corresponding to the historical variables, then they satisfy the two properties of being correlated with the instrumented variable and not with the shocks. It is also possible to use other instruments. For instance Combes et al. (2010) propose to complement historical variables with instruments based on the geological nature of soils. They could have been a determinant of very early human settlements and they have even less chance of being correlated with current productivity shocks.

The econometric literature shows that using IV estimates can create more bias than it eliminates when the instruments are bad. In particular, the instruments must themselves be exogenous, i.e. not correlated with the shocks, and this may be difficult to obtain. The exogeneity of the instruments can be tested when one has more instruments than variables to instrument, using what is called "over-identification" tests. Basically, these tests assume that a sub-group of instruments is exogenous and test whether, under this assumption, the rest of instruments are exogenous. Clearly, when all instruments belong to the same family and are highly correlated between themselves, such tests are not very convincing. Imagine, for instance, that current density is instrumented by the densities of 200 and 220 years ago. Showing that, under the assumption that density 200 years ago is not correlated with current shocks, density 220 years ago is not correlated with current shocks either, and vice versa, is not very meaningful. Consequently, the econometric literature proposes to complement over-identification tests with "weak instruments" tests. They test whether, when one adds a further instrument, this new instrument is sufficiently orthogonal to the others to really provide additional information for predicting the instrumented variable. If the answer is positive, then assuming its exogeneity to test the exogeneity of other instruments will probably lead to a more consistent answer than when it is not. Unfortunately, not all studies estimating agglomeration effects seriously tackle the possible problems of endogeneity; they do not all over-identify their model, and rare are those that propose both over-identification and weak instruments tests.

There are alternative ways to address endogeneity issues. One consists in using natural experiments

that affect the explanatory variables exogenously to identify their causal effect on the local economic outcome studied. For instance, Hanson (1996) or Hanson (1997) uses Mexico's trade liberalisation to such a purpose, and Redding and Sturm (2008) use the division of Germany after the Second World War and its reunification. The problem is that, by definition, the estimates obtained correspond to the natural experiment considered and one knows little about its generality, i.e. the extent to which it is representative of what would occur in another situation generating the same change in the explanatory variable. More frequent is the use of General Method of Moments (GMM) estimates, following Arellano and Bond (1991). This requires time series in the data and is based on IV methods where the instruments consist in lagged levels and lagged differences of the instrumented variables themselves. Henderson (1997) and Henderson (2003) were among the first studies to use this strategy to quantify the local determinants of US employment and productivity respectively. There are two important issues with such methods. First, one cannot give a structural interpretation of the effects of the instruments. It is in general very difficult to get the economic intuition of why they should be valid and in which direction they should modify the estimates. Second, weak instruments tests for GMM methods do not exist yet. Over-identification tests for GMM are based on instruments that all belong to the same family, all the more so because the time-span of data sets used in economic geography studies is in general very short, around 10-12 years at best. Therefore these tests can easily perform very badly and provide little information about the validity of the instruments. On the other hand, it is clear that a GMM strategy makes the search for instruments much easier. In particular, when either the dependent variable or the explanatory variables are sector-specific, long lags or geology do not in general allow for correct instrumentation, since the instruments are not sector-specific and therefore not very effective in predicting the local industrial size, for instance. Another advantage of GMM is that they can be applied to vectors of dependent variables, either to study simultaneously different local outcomes, as Combes et al. (2004) do for employment per firm and the number of firms in the region-industry, or, as Graham et al. (2010) do, to evaluate simultaneously both the causality from inter- and intra-industry externalities to productivity and the reverse causal relationship from productivity to these variables.

For the sake of pedagogy, we have drawn mainly on the link between productivity and density to illustrate this section. However, it should be kept in mind that because of their origins in missing variables and endogenous location choices, endogeneity concerns can affect all the local economic outcomes that are taken as dependent variables and all the effects that are studied in this literature. The effect of local economic size not only on productivity but also on employment, firms' location choices and innovation can be biased when endogeneity is not addressed. Similarly, the effect not only of density but also of market potential, specialisation and the skill composition of the labour force can suffer from similar bias. These variables all result from firms' and households' location choices, which themselves depend on regional productivity, employment and innovation. Again, density is probably the easiest variable to find instruments for, but it can be more difficult when one has to instrument the regional share of a given skill level in a given industry, for instance.

2.2 The magnitude of the effects

The methodological concerns detailed in the previous section may sound a bit boring for policy-makers, but it is important for two reasons. Firstly, this discussion highlights the fact that no definitive answer can be

given concerning the impact of the concentration of economic activities on local economic outcome, because the perfect way of evaluating it has not yet been found. Secondly, care must be taken in interpreting results that may be obtained under very different economic and econometric assumptions. At the same time, there are numerous studies in the field of empirical economic geography, but few of them choose exactly the same way of proceeding in different contexts, and they vary in their degree of achievement. In spite of these caveats, a number of studies deliver interesting and hopefully pretty robust insights. We will now examine these results. Although many regions of the world have benefited from this type of analysis, we focus on the studies on European regions and European countries, since our purpose is to enlighten European regional policy. We only draw on studies of other regions, mostly the US, when they were at the origin of the literature or when they have not yet been replicated in a European context.

2.2.1 Agglomeration economies on productivity and sorting

Density economies

It is now well-known that the local density of economic activities increases the productivity of firms and workers. This conclusion emerges from a large number of studies, of which Rosenthal and Strange (2004) give a comprehensive survey that also covers the role of other local economic characteristics. When one regresses the log of regional wages or of total factor productivity on the log of employment or population density, typical values obtained for the elasticity when using OLS are between 0.05 and 0.09. This implies that when density is doubled, productivity increases by between 3.5 and 6%. Density gaps between regions at the first and third deciles can be as high as a factor of 15 (as, for example, with European NUTS2 regions, but also even within the same country), in which case the productivity gap due to density difference can be as large as 30%.

Only three studies attempt to estimate the productivity elasticity with respect to density for European regions. Ciccone (2002) was the first to replicate Ciccone and Hall (1996), initially designed for US counties, on NUTS3 regions, but European data sets at that time only allowed him to consider France, Germany, Italy, Spain and the UK. The elasticity estimate he obtained was around 0.05. Interestingly, using exactly the same methodology, he found no evidence that agglomeration effects differ significantly between countries. Two more recent studies extended the set of countries considered, although this was at the cost of less geographical detail. Brülhart and Mathys (2008) consider 245 NUTS2 regions of 20 Western and Eastern European countries, spanning the period 19802003 (19902003 for Eastern European countries) and 8 broad sectors covering both manufacturing and financial services. They use a dynamic panel data method (GMM) to deal with endogeneity. Unfortunately, the results they present seem to be pretty dependent on the empirical strategy used to tackle endogeneity (difference versus system GMM). Still, they "confirm the presence of significant agglomeration effects at the aggregate level, with an estimated long-run elasticity of 13%". Interestingly, and possibly consistent with the positive impact on agglomeration of a decrease in trade costs suggested by economic geography models, "repeated cross-section regressions suggest that the strength of agglomeration effects has increased over time". Foster and Stehrer (2009) obtain estimates closer to Ciccone (2002) on a panel of 255 NUTS2 regions in 26 European countries and covering 6 sectors, including agriculture, forestry and fishing that is not considered by Brülhart and Mathys (2008). Their instrumentation strategy, relying on land area as an exogenous instrument only and considering regional skill compositions as exogenous, is not very convincing. Beyond

these limits, they also obtain a further result that is of interest from the perspective of European policies, which is an even larger magnitude of agglomeration economies for new member states than for old ones.

Working on individual countries has the advantage of access to richer data sets than those available at the European level, in particular data sets where individual information may be recorded with a fairly precise knowledge of each worker or firm location. The clear drawback of this strategy is that strictly speaking, the estimates obtained are only valid for a single country. Policy-makers can only hypothesise that similar magnitudes could be obtained in economically similar countries, in particular in terms of those variables (mobility, trade costs and the magnitude of increasing returns to scale) that have been shown by economic geography to be the main determinants of agglomeration effects. This may be reasonable for some European countries but much less so for others. French data allowed Combes, Duranton and Gobillon (2008) to conduct the most complete study, in terms of the treatment of spatial selection and endogeneity, along these lines. Cingano and Schivardi (2004), for Italy, was one of the first studies to use firm-level total factor productivity data, but they did not consider the possible endogeneity of local characteristics. Martin et al. (2011b) do it for the French firms productivity using GMM estimation technics. Italy was the subject of another study by Mion and Naticchioni (2009), who replicate Combes, Duranton and Gobillon (2008) to evaluate the density elasticity of wages while controlling at the individual level for the possible sorting of workers across locations. For the UK, Graham (2007) and Graham et al. (2010) use firm-level total factor productivity data, but only the later instrument the local characteristics variables, using GMM techniques. Earlier studies by Fingleton (2003) and Rice et al. (2006) estimate agglomeration economies in the UK on wages and earnings data aggregated by regions (districts and NUTS3 respectively). Using data aggregated by regions also appears to be the only possible strategy currently available for Spain. Viladecans-Marsal (2004) use a fairly detailed spatial classification considering 331 units to estimate agglomeration economies using GMM techniques. To the best of our knowledge, a large number of European countries, including Germany, have not yet benefited from specific estimates. Melo et al. (2009) propose a meta-analysis that underlines the sensitivity of results to the country studied, the industrial coverage, the way agglomeration economies are specified, and the presence of region fixed effects and controls for the quality of labour. Again, this shows the importance of the strategy adopted when commenting or using a specific estimate. They find that the impact of regional size on productivity is estimated significantly positive in most cases with a magnitude close to 0.05 on average when not instrumented. Variations around this number can deliver interesting further policy insights, which is why we present them now.

The second fairly systematic finding relates to the slight endogeneity bias that is found, due to either aggregate missing variables or reverse causality. Its magnitude is very consistent across studies, at around 20%. Once corrected for endogeneity, the typical density effect of 0.05 would drop to 0.04. This is found both for individual countries, even when instruments of different origins are used as in Combes et al. (2010) for France, and for European regions (Ciccone (2002)). Note that Ciccone and Hall (1996) found almost no difference between OLS and instrumented estimates for the US. This changes the policy implications, as we will see below.

Congestion and industry specific estimates

Theory predicts that too much spatial concentration necessarily leads to congestion. As a result, the effect of density on productivity should be concave. Most studies do not report the estimated degree of concavity;

one can imagine that it is not significant. This sounds surprising, but if believable (it can also come from a more direct lack of variation in the data sets that do not allow identification), it would be crucial for policy. For example, Au and Henderson (2006) for China estimate a bell-shaped relationship between the productivity and size of cities and conclude that most cities lie on the left-hand-side of the peak, i.e. are too small to achieve the highest level of productivity. In Europe, quadratic effects are rarely reported. Martin et al. (2011b) exhibit a positive effect for specialisation that turns negative above a certain level of specialisation. For the UK, Graham (2007) develops an original strategy based on road traffic congestion to estimate the diminishing returns of agglomeration effects and their link with the presence of transport congestion. Five out of nine industries present concave effects of density. Furthermore, it is shown that when congestion is taken into account, the density elasticity increases in seven of the nine industries. This is in line with expectations since, as detailed in section 2.1, in the absence of other controls, the density elasticity reflects the overall net effect of density, taking into account both positive and negative effects. When part of the latter is controlled for, as Graham (2007) does, the density impact must increase. In the UK, congestion is shown to represent up to 30% of the agglomeration effect.

Unfortunately for policy-makers, for whom such information would be quite useful, researchers do not appear to be very industrious in estimating agglomeration effects for each industry separately. One reason may be that the design of the empirical model, and in particular the search for valid instruments, has to be done industry by industry. Another reason is the lack of information for some areas, i.e. the non-availability of data per industry, as in Ciccone (2002) for instance. Brülhart and Mathys (2008) and Foster and Stehrer (2009) are notable exceptions, and they work at the European regional level. They find significant agglomeration effects in all but one of the sectors they consider. The exception is agriculture, on which regional density has a negative impact, which is fairly intuitive. Given the share of land in agricultural production and the fact that land prices increase with density, less dense places clearly represent the best alternative for productivity in this sector. As regards individual countries, Graham (2007) finds larger effects for services than for manufacturing in the UK.

Spatial extent of density effects

The rapid spatial decrease of agglomeration effects is another robust finding in the literature. Agglomeration economies do not spill over regional boundaries so much. The advertising agency industry example proposed by Arzaghi and Henderson (2008), with an extremely fast spatial decay of agglomeration effects that are shown to occur primarily within 500 meters, is certainly too extreme to be representative of more standard industries. Still, effects are rarely found to be significant beyond 100 kilometres, and the threshold is often shorter. The first way to appreciate the role of the spatial diffusion of agglomeration effects consists in considering market potential variables. Some authors include a single market potential variable that consists in the spatially discounted size of markets including the region's own (divided by the region's internal distance). Alternatively, one can introduce two variables separately, the region's own size and the market potential that the other regions represent - the external market potential. Head and Mayer (2006) tests different variants of the first strategy, the total market potential, on NUTS2 European regions. They consider both a reduced-form market potential, which simply consists in the sum of regional GDPs divided by the distance between regions, and more structural ones that correct for price effects. Both kinds of variables are shown to significantly affect regional wages, and share a similar

explanatory power. Importantly, when one divides by distance, measured in kilometres, in the market potential definition, one implicitly assumes a pretty strong spatial decay of agglomeration effects. The impact on a region of the activity located 20 kilometres away is four times lower than the activity at 5 kilometres, it is 10 times lower at 100 kilometres than at 10 kilometres, and so on. And in fact, this functional form is rarely rejected empirically, whether it is estimated at the European level, as in Head and Mayer (2006), or within countries, as it is for instance the case in Holl (forthcoming) for Spain who considers distance through the real road network instead of the 'as-the-crow-flies' distance to compute the market potential variable and who instruments it using both internal and external (historical population, geology and historical transport networks) instruments. External market potential is also significant when this form is used to compute it and when it is introduced next to density in the specification. For instance Combes, Duranton and Gobillon (2008) and Combes et al. (2010) find that both density and external market potential, computed using inverse distances, are significant for France, even when both are instrumented and the regional labour skills composition is controlled for. Foster and Stehrer (2009) for NUTS2 European regions also consider a spatial decay of spillover effects from other regions introduced next to the density effect, but with an exponential form, i.e. assuming an even sharper decline than the inverse function. Not only is this functional form not rejected but, testing various exponential functions, only those with the strongest spatial decay show significant effects. Note that, in general, considering spillover effects beyond regional boundaries barely affects the estimate of own-density impact.

The second strategy for testing the spatial decay of agglomeration economies consists in introducing the size of other regions located at various distances from the own region. For instance, at the NUTS3 European level, Ciccone (2002) finds that production in neighbouring regions also positively impacts own-region productivity. Ciccone (2002) does not give the magnitude of the coefficient however, and he does not test the impact of regions located further away. The first two studies to introduce the size of other regions separately were not aiming to explain local productivity, but local firm creation (Rosenthal and Strange (2003)) and employment (Desmet and Fafchamps (2005)), both on US data. The idea consists in considering concentric rings around the location under study, and introducing their size (in terms of employment, for instance) as explanatory variable. In Rosenthal and Strange (2003), local activity is located within 1 mile of the zip code centroid, and then they consider three rings. The first ring contains activities located between the 1 and 5 mile circles, the second between 5 and 10 miles, and the third between 10 and 15 miles. In Desmet and Fafchamps (2005), the first ring corresponds to activities located between 0 and 5 kilometres from the county, between 5 and 10 kilometres for the second ring, then 10 to 20 kilometres and so on every 10 kilometres up to 100 kilometres. The results of these studies are detailed in the next section on employment and firm creation. As regards productivity, Rice et al. (2006) find for the UK that agglomeration externalities attenuate sharply with distance. Distant markets do affect local wages and productivity, but the effect from those located between 40 and 80 min travelling time is four times lower than from those located at less than 40 min, and the 80-120 minute markets have an effect twice as low as the 40-80 minute ones. Basically, there are no effects beyond 80 minutes. Rosenthal and Strange (2008) obtain even larger spatial gradients on wages in US cities. The effect on the 0 to 5-mile variable is four to five times larger than on the corresponding coefficient on the 5 to 25-mile variable. Turning to the outer rings (25 to 50 miles and 50 to 100 miles), effects are even smaller and very often not significantly different from zero.

Specialisation and diversity

An old question that Glaeser et al. (1992) brought up to date relates to the fact that the size of the overall local economy matters for workers' and firms' productivity or for the local size of the industry in which the worker or firm operates. Initially, many authors included both the local level of total activity and the local level of industry activity. As Combes (2000) shows, this strategy raises a serious identification concern since the two variables are strongly correlated. When overall local markets are large, industrial markets are large too. What one wants to capture is a composition effect and the role of specialisation. Therefore, next to the effect of the size or density of total activity, the share of the industry in the local economy, and not its local level, must be introduced into the specification. As an example on European data, this is the own industry density that Brülhart and Mathys (2008) consider in their specification, but it is difficult to assess whether this variable captures an effect of the industry or of the total employment density (typically when introduced simultaneously, the density in other industries is not found significant, whereas it certainly would be when introduced alone). In many studies, when both total density and specialisation are simultaneously introduced, both are found to significantly affect productivity. For instance Cingano and Schivardi (2004) show that this is the case in Italy when industries are pooled together. They also find that the spatial decrease is very strong, since specialisation in neighbouring regions has no impact on local productivity. Martin et al. (2011b) find significant positive effect of specialisation on firms productivity in France. Still for France, Combes, Duranton and Gobillon (2008) find that the effect of specialisation, estimated separately for each industry, is significant in 94 industries out of 99 and that its magnitude is larger in business services and in two high-tech industries - medical instruments and artificial fibres. This is intuitive and confirms the findings of Henderson (2003) for the US of a larger effect in high-tech industries. Interestingly, Combes, Duranton and Gobillon (2008) also show that if total employment density explains a large share of the spatial disparities in productivity, that is not the case for specialisation. This gives the two variables different implications in terms of optimal policy.

Following both the intuition of Jacobs (1969) and the central role of the preference for diversity in many economic geography models, another appealing variable to explain productivity is the overall industrial diversity of the location. However, its effect has been shown to be not robust. It is sometimes found to be significantly positive, sometimes significantly negative, and often not significant at all, as for example for France in both Combes, Duranton and Gobillon (2008) and Combes et al. (2010), for Italy in Cingano and Schivardi (2004) and for the US in Henderson (2003). Beyond the nice intuitions behind these variables, the empirical effect does not really seem to be present. It may be due to the way diversity is assessed, which is often based on Herfindhal indexes computed over the shares of each industry in the local economy, but using data that are often quite aggregated. Some industries may use a group of other industries but they will rarely draw on all of them as assumed in the Herfindal indexes. To tackle this issue, Moretti (2004b) uses a measure of proximity between industries and finds for the US that spillovers between economically-close industries are larger than spillovers between economically-distant industries, but that sounds rather trivial now. The correct way to test the role of diversity has not really been found yet.

Locals skills

Another important concern relates to the role of high-skilled workers and the fact that they can exert a positive localised externality on other workers. Once again, this raises an identification issue. When the average local productivity is regressed on the local share of professionals or highly-educated workers in the region, it captures two effects: the possible externality effect of these workers but also, and more directly, the fact that they are themselves more productive than other workers. Using aggregate data, it is not possible to separate the two. On the contrary, when using individual data, one can introduce into the wage equation both the own-skill level of the worker and the skill level of the other workers in the same region, and thus identify separately the role of own-skills and the externality effect. Still, according to Ciccone and Peri (2006), there is an interpretational issue due to the fact that the standard wage approach confuses positive externalities with wage effects due to a downward sloping demand curve for human capital. The standard wage approach has been proposed for the US by Moretti (2004b), whereas Ciccone and Peri (2006) attempt to identify the two effects of high skills separately, and indeed the two obtain contrasting results. While the former conclude that there is a positive externality effect of college graduates on productivity, the latter conclude that there are no significant average schooling externalities. With this important remark in mind, Rosenthal and Strange (2008), who use the standard wage approach, find the same positive role of college-educated workers in the US. Considering the variable at various distances from each worker location as they do for density, they also obtain that human capital effects attenuate sharply with distance. The 0 to 5-mile estimate for proximity to college-educated workers is 3.5 times larger than the corresponding 5 to 25-mile effect. In Europe, Rice et al. (2006) control for the local share of degree-level qualification and find that it has a positive effect on wages and productivity. However, since the specification is not estimated at the individual level, it is impossible to assess the composition and externality effects separately. This is possible for France, and even controlling for an individual fixed effect and the role of gender and age that capture fairly precisely the own-skills of each worker, Combes, Duranton and Gobillon (2008) find a positive externality emanating from the professionals located in the same employment area as the worker.

Interestingly, when both productivity and wage data are available, one can evaluate how much of the productivity gains due to agglomeration are transformed into wage gains for workers. Unfortunately, this has not been done for Europe, but for the US Moretti (2004b) finds that estimated productivity differences between cities with high human capital and low human capital are similar to observed differences in the wages of manufacturing workers, indicating an almost complete offset. Rice et al. (2006) decompose regional average earnings in the UK into a productivity index and an occupational composition index. They find that about two-thirds of the spatial variance in earnings can be attributed to variations in productivity.

Individual data allow researchers to identify the effects of density and individual skills separately. Combes, Duranton and Gobillon (2008) show on a very large French data set that individual characteristics explain more than half of wage disparities. They are able to control not only for age and gender but also for an individual fixed effect, which is identified from "movers" and "stayers" next to an area fixed effect. The elasticity of productivity with respect to density is much affected by the consideration of such individual fixed effects. It is divided by two even by comparison with aggregate regressions that do control for aggregate human capital. Instead of being at 0.04 (once reverse causality has been taken into account),

it is only at around 0.02. Combes, Duranton and Gobillon (2008) also show that individual abilities do not distribute randomly across locations. Higher-skilled workers, even in terms of non-observable characteristics, locate more to denser cities, the correlation between individual and area fixed effects being as high as 0.29. The correlation between individual fixed effects and density is 0.44. This is the fundamental reason why controlling for individual characteristics has so much influence on the estimate of the productivity elasticity with respect to density. Mion and Naticchioni (2009), who replicate the study on Italian data, obtain similar conclusions, although to a lower extent. The correlation between individual fixed effects and density is still significantly positive, but only at 0.21. The decrease in the density elasticity remains large and the density effect net of individual sorting is tiny: the elasticity decreases from 0.022 to 0.007. Clearly, this interacting role of individual and location characteristics has many policy implications, which we discuss below. It may, for instance, have a strong impact on the time evolution of regional earnings disparities. As an example, Duranton and Monastiriotis (2002) show that if education returns and their distribution across the UK had remained stable over the 1982-1997 period, the UK North-South income divide would have fallen, whereas it actually increased quite sharply.

Lastly, a few studies have sought to evaluate the extent to which agglomeration economies can be stronger for some types of workers or firms. For instance, Bacolod et al. (2009b) confirm the intuition that returns to education are higher for high skills in cities. In the same vein, but for firms, Combes et al. (forthcoming) obtain that the most efficient firms gain more from density than the least efficient ones. For instance, firms in the first quartile of productivity gain 3 times less from density than those in the last quartile. The authors also find that the largest establishments gain more from density, 50% more for those with more than 100 employees compared to those with 6 to 10 workers. Going in the opposite direction, Henderson (2003) and Martin et al. (2011b) conclude that specialisation effects are larger for smaller firms. Other authors have sought to investigate the source of the gains from agglomeration. For instance, Rosenthal and Strange (2003) on US data find that the number of hours worked decreases with density for non-professionals but increases for professionals, and the effect is stronger for young workers. Moreover, the number of hours worked by young professionals is particularly sensitive to the proximity of other young professionals. These patterns are consistent with the presence of hard workers in cities and a higher productivity of agglomerated labour that would lead to larger individual effects there. Bacolod et al. (2009a) consider which sorts of skills have returns that are positively related to city size. They conclude that only cognitive and social skills are better rewarded in large cities, while motor skills and physical strength are less-well rewarded. Using an original data set of workplace communication practices in France, Charlot and Duranton (2004) find that in large and educated cities, workers communicate more and this has a positive effect on their wages. This explains 13 to 22% of the effects of a better-educated and larger city on wages. Unfortunately, these studies that shed some light on the microeconomics and channels of agglomeration effects remain to be generalised and replicated in other European regional contexts. As an example, we still know very little about the sources of agglomeration effects, even though this would greatly extend the policy implications that can be drawn from empirical economic geography studies.

2.2.2 The dynamics of regional employment and firm creation

The literature on the influence of the characteristics of the local economy on the local economic outcome was revived through the studies of Glaeser et al. (1992) and Henderson et al. (1995). However, these authors were not working on productivity as the dependent variable, but employment growth. They made the implicit assumption that positive effects on employment growth result from positive effects on productivity and interpret their results accordingly. However, we have stressed above that positive productivity effects only result in positive employment growth effects under certain conditions on the supply and demand elasticities of labour. The main reason why these authors, and many others, work on employment rather than productivity probably relates to the fact that total factor productivity or even wages per industry were initially less available in data sets at fine geographical levels such as cities or labour market areas. Research on European economies faced the same trade-off and the first studies for this continent also often estimated the local determinants of employment growth only. Unfortunately, the fact that a positive effect on local productivity can lead to a negative effect on employment was confirmed by Cingano and Schivardi (2004) on Italian data. They get opposite signs for some of the common determinants of productivity and employment growth, based on the same data set. As explained above, this implies that research focusing on only one of these two components can be frustrating to interpret. Studying each separately can be useful for policy-makers, but it does not allow one to infer very much about the other component or the underlying effects. We must bear in mind that enhancing productivity or employment may not necessarily call for the same tools, and that a policy enhancing one may be detrimental to the other.

Another concern with employment growth regressions is that one has to control in the right-hand-side for wages and production, bearing in mind that these variables are highly endogenous and therefore need to be carefully instrumented. Otherwise, the interpretation of the effects changes. Apparently, not all authors were aware of these caveats. Still, a number of results have been obtained, which we will now detail.

Total employment, specialisation and diversity

The explanatory variables that are introduced into employment growth regressions are very similar to those considered in productivity regressions. The impact of total employment on industrial employment growth is a first good example of the variety of results obtained in this literature. For instance, Combes (2000) finds for France that total local size is growth-enhancing for manufacturing sectors but detrimental to growth in service industries. Viladecans-Marsal (2004) for Spain finds it non-significant in three out of six sectors, while it has a bell-shaped effect in the three others. Blien et al. (2006), who extend Blien and Suedekum (2005), find results for Germany that are broadly speaking very consistent with those obtained for France. Still, as regards the impact of total local size, it plays a positive role on industrial employment growth for both manufacturing and services activities. There are two recent studies on Italy, one that pools together both manufacturing and service industries (Mameli et al. (2008)) and the other that concentrates on business services (Micucci and Giacinto (2009)). Both conclude that total employment has a positive impact on industrial employment.

As we mentioned above, the question of the spatial decay of agglomeration effects is crucial for regional policy. For the US, Desmet and Fafchamps (2005) show that for the non-service sectors such as

manufacturing and construction employment has moved away from centres of high aggregate employment to nearby locations. The coefficients are negative for distances below 20 kilometres, and are then slightly positive for distances between 20 and 70 kilometres. Service sectors exhibit a different pattern: they grew faster in aggregate clusters and slower in nearby areas. The coefficients are positive at distances below 5 kilometres, and slightly negative at distances between 5 and 20 kilometres. Unfortunately, this question has rarely been addressed for European economies. Viladecans-Marsal (2004) studies the role of the local characteristics of neighbouring cities in Spain. She finds it to be significant in three out of the six sectors considered. In the same vein, and still on Spanish data, Solé-Ollé and Viladecans-Marsal (2004) show that growth of the central municipality (within metropolitan areas), whether it is measured in terms of population or economic activity, has a positive growth effect on the suburbs. Micucci and Giacinto (2009) on Italy also find a significant impact of distant locations.

The impact of diversity was found not to be very robust on productivity. On industrial employment growth it is also very variable from one context to another. According to Combes (2000) for France, it has a positive impact in services and a negative impact in most manufacturing industries, although there are a few for which it is positive. Viladecans-Marsal (2004) finds it significantly positive in three sectors, but significantly negative in another and non-significant in the last two. For Germany, Blien et al. (2006), using a slightly different diversity index, find that it has a positive effect on employment growth in both manufacturing and services, but that it is stronger in the former. Diversity is also found to have a positive significant impact in Italy, according to Mameli et al. (2008).

The role of specialisation is difficult to assess because it cannot be disentangled from the mean reversion process. The specialisation variable - the share of industrial employment in total employment - is directly correlated to initial industrial employment, which captures mean reversion, and total employment, which estimates the role of the local economy size. It was initially suggested that the effect of local specialisation could be identified using a non-linear effect of industry employment that enters the specification both in logarithms and levels. However, this makes interpretation difficult when the two effects act in opposite directions, as for instance in Henderson et al. (1995), and it is better to avoid that. Therefore, the negative impact of specialisation, which is found to be negative for France by Combes (2000) in both manufacturing and services or for Italy by Mameli et al. (2008), may only be due to a strong mean reversion rather than compensating positive specialisation effects. Viladecans-Marsal (2004) finds for Spain a significantly positive impact of specialisation in two industries. Blien et al. (2006), estimating a full dynamic model for industrial employment, observe a slight tendency of mean reversion, which would be consistent with the co-existence of mean reversion and positive specialisation effects. But the latter would not be strong enough to reinforce growth permanently and lead to an explosive dynamics to create a full concentration of industrial employment in a single area (for each industry).

Glaeser et al. (1992) popularised the use of a variable corresponding to the average size of firms in local industry. The interpretation they gave to it, in terms of competition, was particularly misleading, since firms located in a given location compete with firms located in all other locations and in many different markets (for both outputs and inputs). This makes the notion of competition difficult to assess, since we do not know on which markets it applies, in terms of either locations or goods. Moreover, the direct interpretation in terms of firms' size is interesting in itself, since it might reflect the impact of internal increasing returns to scale that economic geography models emphasise. Both Combes (2000) for France and Blien et al. (2006) for Germany find that the presence of larger firms reduces employment growth in

both manufacturing and services. To capture the role of the local firm size, Combes (2000) introduces a local Herfindhal index of firm size heterogeneity. He finds that this variable is also detrimental to growth. Therefore, the structure that most favours employment growth in France is composed of small firms of even size. Mameli et al. (2008) illustrate for Italy that the effects of these variables might not be very robust, in the sense that their sign changes depending on the industrial classification used. They turn positive at the three-digit level while they are negative at the two-digit level.

Dynamic models

A crucial question for regional policy regards the time frame during which a variable has an impact. The higher the speed of adjustment, the faster the impact of the policy, but also the less long-lasting its effects. The availability of time series in regional data sets has generated a series of papers that estimate the dynamics of the variable studied (industrial employment for instance) alongside the dynamic effects of other local characteristics. In other words, instead of simply studying the determinants of local growth, researchers estimate full auto-regressive models, as Henderson (1997) proposed initially for US cities. Interestingly, this also allows them to address endogeneity concerns using GMM estimation techniques. Moreover, short-run effects can be distinguished from long-run effects. For instance, Blien et al. (2006) for Germany show that the impact of diversity they had observed dies out quickly over time, both for manufacturing and services. This means that it had no long-run effects. Similarly, the firm size effect, significant in the short-run, has no significant long-run impact on either manufacturing or services.

Combes et al. (2004) extend Henderson (1997) by decomposing industrial employment into average employment per firm and the number of firms in the local industry and then estimate a Vector Auto-Regressive model. From the policy point of view, this allows them to determine whether local characteristics influence the two components of local growth in the same way. These two components are internal growth or the intensive margin, which concerns the growth in size of existing firms, and external growth or the extensive margin, which concerns growth in the local number of firms. They find that some effects can indeed be different for the two, which brings to light the presence of a trade-off between these two possible sources of growth when one wants to boost local employment. The elements conducive to the growth of existing firms are not necessarily the same as those that promote the creation of new firms. More precisely, it appears that a large number of different-sized plants positively influences the growth of existing plants, whereas more new plants tend to be created where there are a small number of plants of a similar size. A large regional labour market with a small number of similar-sized industries would favour the growth of both new and existing firms. Another conclusion of this study on France is that the adjustment process shows greater inertia in the United States than in France, since lagged values stop being significant after one year. It is starkly at odds with the six- or seven-year significant lags found in Henderson (1997). Moreover, whereas area-and-industry effects explain most of the spatial variation in plant size, the local number of plants is mainly driven by the current local economic structure. Policies targeting plant creation should thus be more efficient. Fuchs (2011) replicated the study by Combes et al. (2004) on German data. They emphasise the positive influence of diversity on both the intensive and the extensive margin, whereas there is no clear result for specialisation. Plants of similar size promote growth in the number of firms but they are detrimental to firm size, as in France. Hence, these dynamic panel regressions show that static externalities dominate again, as in France and contrary to the US.

Importantly, Fuchs (2011) also show that the impact of the local industrial structure on employment dynamics does not differ between small and larger plants.

In the 1990s, a number of papers studied the determinants of regional unemployment - an important issue for the European Union given the magnitude of disparities in unemployment rates. Surprisingly enough, we found almost nothing relating this early literature to the recent developments in empirical economic geography. For instance, we found one paper by Jurajda and Terrell (2009) that studies the role of human capital for local unemployment in some Eastern European countries (Czech Republic, Hungary, Ukraine), but clearly they do not extend their analysis to the role of overall density, specialisation or diversity, and their concerns about possible endogeneity issues due to endogenous location choices are really minor.

2.2.3 Firms location determinants

We have seen that some authors extend their analysis of the determinants of local employment to the determinants of the local number of firms. This number is the net result of the destruction and creation of firms. Arguably, focusing on the latter and, by extension, on the location choices of new firms, can also shed light on the magnitude of agglomeration effects. A specific literature has been devoted to this question, which we present in this section.

Strategies and methodological concerns

There are various possible empirical strategies to evaluate the determinants of location choices. The most appealing consists in applying discrete choice models, as developed by McFadden (1974), to firms' location choices, as first implemented by two early studies, Head et al. (1999) for the US and Guimaraes et al. (2000) for Portugal. The main limits to this strategy concern the relatively small number of possible choices that can be considered, even if the increase in computing capacities allows researchers to consider frameworks encompassing more and more of them. As described above, a first alternative to reducing the number of possible choices consists in specifying nested logit models. By imposing a timing structure, decisions are made in two or three steps, for instance. The number of alternatives is reduced at each step, which makes the estimation possible. However, it must be noted that the sequentiality of choices is important. Another alternative adopted in a number of papers consists in taking as the dependent variable the number of location choices made in a region and then using count models such as the Poisson or the negative binomial models, or simply a Tobit approach. This latter strategy only corrects for the fact that the dependent variable is left-censored but then treated as a continuous variable. The main advantage with count models is that there is no longer any limit on the number of alternatives. However, these empirical models may make strong distributional assumptions for their residuals. All these possibilities have been considered in empirical studies.

From the more economic perspective, a clear distinction has to be made between the location choices of foreign firms (the largest body of contributions, relating to the literature on FDI - Foreign Direct Investment) and the location choices made by national firms. Then, we set aside the role of regional policies, taxes and public infrastructure (which will be presented in a separate chapter below), the variables used to explain the location choices are very similar to those considered in the literature on the determinants of local productivity and growth presented in previous sections. Whereas the dependent variable

often relates to individual location choices, the scale at which agglomeration effects are studied varies considerable between studies. For instance, in the two studies we have just mentioned, Head et al. (1999) consider agglomeration effects within the 50 states of the US, while Guimaraes et al. (2000) consider 275 "concelhos" in Portugal, which are clearly much smaller in size. Because of the regional perspective of this report, we shall not discuss studies based on location choices at the country level. It can be noted, however, that their findings do not significantly differ from those found at the regional level. Interested readers are referred to two interesting recent studies adopting this country perspective: Disdier and Mayer (2004), who compare agglomeration economies for inward FDI from French multinationals in Eastern and Western Europe and Buch et al. (2006), who examine FDI determinants of German multinational firms in various host Central and Eastern European countries.

Once again, if we accept that endogenous location choices make almost all of the explanatory variables endogenous, this raises a number of endogeneity concerns for most of the estimation techniques, as we explained in section 2.1. Unfortunately, these issues are tackled even less often for location choices than in the literature on productivity and growth determinants. At best, authors lag variables by one period of time, which is certainly not enough to correct any endogeneity bias that may be present. Addressing the question of possibly omitted regional variables, some authors include regional dummies either at a geographical scale larger than the one considered for agglomeration effects (8 regional dummies for the US states in Head et al. (1999) for instance, 2 dummies for the concelhos in the districts of Lisbon and Porto in Guimaraes et al. (2000)), while others exploit the presence of time series to introduce a fixed effect for each location (as Hilber and Voicu (2010) do for Romania, at the regional NUTS2 level). This presupposes that there are enough time variations in agglomeration effects to identify them only over time variations, which is not at all certain, given the inertia over time of the regional hierarchy and characteristics. For instance, this lack of time variability could explain why many regional characteristics are no longer significant when fixed effects are considered by Hilber and Voicu (2010). However, this result could also be due to the real presence of omitted variables driving both regional characteristics and location choices, such as a suitable geographical environment or pre-existing public infrastructure of transport or education.

Local economy size and market potential

The first factor that is almost systematically found to play a positive role on location choices of FDI is, as predicted by theory, the size of the local economy. For instance, Head et al. (1999) proxy it by local income, and Guimaraes et al. (2000) by two variables on manufacturing and services employment respectively. The latter also control by the distance to the main cities of Portugal, but this kind of variable has subsequently become widespread with the use, as for productivity equations, of market potential variables. For instance, at the European level, Head and Mayer (2004) compare the performance of simple and structural market potential variables in explaining the location choices of Japanese affiliates across European regions at the NUTS2 level. They find that both of them have a significantly impact on these choices, even controlling for a substantial number of other variables. Two other studies on European regions by Basile et al. (2008) and Basile et al. (2009) analyse the location choices of multinational firms of different nationalities in 50 (47 respectively) regions of eight (five respectively) EU countries (France, Germany, Italy, Spain, Ireland, the UK, Portugal and Sweden, the last three not being considered in Basile et al. (2009)). Simple market

potential is also found to be significant in these studies, again even when it is introduced next to an overall agglomeration variable that also positively affects them. However, it is also shown that the effect is mainly driven by a large impact for European multinationals, the effect being non-significant for non-European ones.

As in the literature on productivity determinants, the functional form chosen for the impact of distance in the market potential function - the inverse of distance in most cases - implies a fast spatial decay of agglomeration effects. Such a fast decay is also found by Rosenthal and Strange (2003) for the creation of establishments (national or not) in the US using a more flexible form that estimates the impact of activity at various distances separately. The initial attenuation is rapid, with the effect of own-industry employment in the first mile that is from 10 to 1000 (depending on the specification) times larger than the effect 2 to 5 miles away. Beyond 5 miles, attenuation is less pronounced. At a lower geographical scale, Arauzo-Carod and Viladecans-Marsal (2009) for Spain show that the within-city spatial decay of agglomeration effects for new firm creation is also quite strong, and increasing with the technological level of the industry. Basile (2004), for FDI in Italy, even find a negative effect of adjacent-province agglomeration, while at the same time the own-province agglomeration has a positive effect. Interestingly, Basile (2004) can distinguish foreign acquisitions and greenfield investments. The effect of the local number of establishments is found to be significantly positive only for the former. However, local demand measured by electricity consumption, introduced into the specification at the same time, exerts a positive influence on both. Greenfield investments are more appealing for evaluating the role of agglomeration effects, because in this case firms have more discretion in their location choices. The impact of market potential seems to be fairly universal and it is confirmed when data is disaggregated along various dimensions. For instance, Crozet et al. (2004) show that it is positive on FDI in France whatever the country of origin of the firm. Spies (2010) studying FDI in Germany also finds it to be positive for all sectors when estimated for each separately. A recent study by Pusterla and Resmini (2007) focuses on FDI in the NUTS2 regions of four Eastern European countries: Bulgaria, Hungary, Poland and Romania. Both local manufacturing employment and market potential variables positively affect FDI, although most of the impact is on low-tech industries, when they are distinguished from high-tech ones.

Past foreign presence

This literature almost systematically considers the role of a variable absent from local productivity or growth estimations: the magnitude of past foreign presence in the region. This variable has two main interpretations. On the one hand, it may act positively on future FDI because it reflects unobservable characteristics of the region that are also beneficial to new FDI, or because it is related to the role of business networks that may exist between foreign firms. One the other hand, it may have a negative impact on new location choices by a simple competition effect, but this appears to be systematically dominated. For instance, a positive effect of past FDI is found in Head and Mayer (2004) for the location choices of Japanese affiliates in European regions, in Italy (Basile (2004)) for both acquisition and greenfield investments and also in adjacent provinces for acquisition, in Germany for all sectors (Spies (2010)) and both low- and high-tech industries, in Eastern European countries (Pusterla and Resmini (2007)) and in Ireland (Barrios et al. (2006)). This is slightly qualified by Basile et al. (2009), who estimate a positive effect for the impact of foreign presence on European FDI but not on non-European FDI (although Basile

et al. (2008) find it positively significant for both). Crozet et al. (2004) find it relevant in France for FDI from some countries only, the largest effects being observed for Japan, the UK, Belgium, and the US. Finally, Devereux et al. (2007), the main aim of which is to evaluate the role of regional assistance (as detailed in the next chapter), find a positive effect of past foreign investment on both new investment by domestic firms and FDI, the effect being larger for the latter. The role of potential social and business network effects has also been indirectly investigated through the effects of variables like the distance to home country or to headquarters, which is found to have a negative impact on FDI in France by Crozet et al. (2004) and in Europe for European FDI by Basile et al. (2008). Generally, sharing a common language also has the expected positive effect on FDI.

Industrial composition

In the same spirit as productivity determinants, researchers also study the effect of the industrial composition of the local economy on location choices, with the same priors as for productivity. Specialisation, typically the domestic industry count, is fairly systematically found to exert a positive influence on FDI location choices, as much at the European level (Head and Mayer (2004)), as for individual countries (for instance Portugal in Guimaraes et al. (2000) or France in Crozet et al. (2004)). This positive impact is also found on domestic firm creation, for instance in Spain by Arauzo-Carod and Viladecans-Marsal (2009) at the city level, which confirms the result of Costa-Campi et al. (2004) at the Spanish inter-regional level, or in the UK by Devereux et al. (2007), the effect increasing with the spatial concentration of the industry. Devereux et al. (2007) also estimate a positive impact of industrial diversity, while for Spain the impact of diversity on firm creation depends on the sector studied (Arauzo-Carod and Viladecans-Marsal (2009)). For Ireland, Barrios et al. (2006) find that diversity has had a significantly positive impact on FDI since the 1980s, but not before, and only for high-tech firms, for which specialisation has no impact. Conversely, whereas diversity does not matter for low-tech firms, specialisation has a positive impact on their location choices. Hilber and Voicu (2010) for Romania find that both domestic and foreign industry-specific agglomeration positively affect FDI, but only the former is robust to the introduction of regional fixed effects. The same is found for domestic industry-specific agglomeration in neighbouring regions. The positive effect of diversity that is estimated without regional fixed effects is found to be not robust to their introduction.

We have already noted that Guimaraes et al. (2000) distinguish between the roles of manufacturing and services concentration and study their impact separately, finding the effect of services to be stronger. The prominent role of services has been confirmed in later studies, and for Eastern European regions particularly. According to Cieślik (2005), it has a significantly positive and large effect on FDI in Poland at the NUTS3 level (49 regions studied), and the same is found for Romania at the NUTS3 level (21 regions) by Hilber and Voicu (2010), even when region fixed effects are included in the specification. As an example, an increase of 10.0% in service employment density in a Romanian region makes the average Romanian region 11.9% more likely to attract a foreign investor.

Regional labour market characteristics

Another variable that is much studied in this literature is local labour costs. Unfortunately, it difficult to interpret because authors are rarely able to control simultaneously for the local quality of labour.

Typically, this is clearly the cost per efficient unit of labour that theory would predict to influence location choices. When labour productivity is not controlled for, a positive impact of wages on location choices may simply reflect the presence of highly-skilled workers. Similarly, some papers introduce regional unemployment rates into the specification, but again many counter-effects arise. Higher unemployment rates may reflect a large local supply of labour, hence low wages, or on the contrary, too high a wage if this is what causes unemployment. When wages are controlled for, the unemployment rate can mainly reflect a depressed area and low demand, if this is not properly controlled for elsewhere in the specification. At the European level, both Head and Mayer (2004) and Basile et al. (2009) find no significant effects of either regional wage or unemployment (unemployment only has a significant negative impact for EU FDI in Basile et al. (2008)). This result can be clearly explained by the caveats we have just mentioned; although education is controlled for in both Basile et al. (2008) and Basile et al. (2009), it does not have any significant impact either. Crozet et al. (2004) also find no effect of unemployment in France for average FDI, but when estimated separately for each country source of FDI, it is found to negatively affect investments from some of them (Belgium, the Netherlands, Italy), and to positively affect them for FDI from the US. For other countries, unemployment is non-significant. The effects of unemployment and high-school degrees are not robust to the introduction of regional fixed effects in Romania, as shown by Hilber and Voicu (2010), while labour conflicts are never significant in his specification. Similarly, for Poland, Cieślik (2005) finds no robust effect of regional wages and education, while unemployment exerts a negative effect. Basile (2004) finds that both unit labour costs (which control indirectly for productivity) and unemployment have a significantly negative impact on FDI in Italy for both acquisition and greenfield investments. In Germany, Spies (2010) estimate it to be significant and negative in only one sector ("other services") out of four. In Pusterla and Resmini (2007), who focus on NUTS2 regions in Eastern European countries, wages, when controlled for education, only have a negative impact on low-tech sectors. Finally, for the UK, Devereux et al. (2007), who distinguish between low-skilled and high-skilled wages find a negative effect for the former and a positive effect for the latter, and a slightly significant negative effect of unemployment.

Entrepreneurs' location choices

Finally, beyond new firms' location choices, one can argue that entrepreneurs' location choices and their determinants could also be interesting to study and could shed light on the role and magnitude of agglomeration effects. Unfortunately, as Glaeser et al. (2010) observed, the economic geography literature on this topic is relatively small. Few studies are devoted to European countries, beyond the couple of papers quoted above, which study the location choices of new firms more generally (Arauzo-Carod and Viladecans-Marsal (2009), Devereux et al. (2007)). One can mention Figueiredo et al. (2002), who analyse the location choices of entrepreneurs in Portugal. They find that, as in the case of firms' investments, the prior base of economic activity is the most important driver of location choices. Interestingly, these authors are able to distinguish between home and non-home (of the entrepreneur) location choices. They show that whereas for the non-home choices, agglomeration effects of the kind we discussed for FDI are present, this is not the case for home choices.

2.2.4 The spatial diffusion of knowledge spillovers

We turn now to those studies that focus on a specific mechanism behind agglomeration effects, namely the imperfect diffusion across space of knowledge and innovations, and hence the presence of localised "technological spillovers", as they are called in the literature.

Knowledge production functions and spillovers in the US

Jaffe (1989) is considered as one of the first studies proving econometrically the causal role of proximity on innovation. He studies how the patenting activity of private firms relates to the R&D done in universities located nearby, controlling for the firms' own R&D and a number of reverse causality issues. Typically, he shows that research done in universities located in the same US states as the firms has a positive significant impact on private firms' patenting. There is however no significant impact of co-location within the state. It has been correctly observed that patenting does not necessarily imply innovation. Some further studies, such as Acs et al. (1992), have used variables from innovation surveys to confirm this role of proximity on innovation. Explaining innovations, the conclusions of Jaffe (1989) are not only confirmed but reinforced. The impact of university research on innovation is larger than on patenting, and co-location also has a significantly positive effect.

As for the impact of local characteristics on growth, a big issue relates to the spatial extent of the agglomeration spillovers beyond co-location in the same region. Adams and Jaffe (1996) and Adams (2002) analyse innovation in the same way as impact of density, by introducing into the specification some variables computed at various distances from the location studied and comparing the magnitude of their respective effects. The former estimate the impact of research over 100 miles to be only 20% of the effect within 100 miles. The latter obtain significant effects within 200 miles only, and academic spillovers appear to be more localised than industrial spillovers. Spillovers are also found to be weaker and even more localised for new products than for patents. Similar conclusions are reached by Anselin et al. (1997) working at the level of US metropolitan areas and using slightly different measures of spillovers. Another question this literature tackles relates to the interaction between the size of firms and their ability to benefit more or less from technological spillovers. According to Acs et al. (1994), all firms benefit from spillovers, but large ones benefit more from private investment while small ones benefit more from university / public R&D. Finally, the role of local industrial composition in terms of specialisation and diversity on innovation has been investigated too, for instance by Feldman and Audretsch (1999) on US data. While they find a significantly positive role of diversity, specialisation does not seem to increase the innovation rate. Still for the US, Carlino et al. (2007) show that patenting is positively related to the density of employment. A city that is twice as dense exhibits a rate of patents per capita that is 20% higher, the effect being concave.

Knowledge spillovers in Europe

Some authors have attempted to replicate these findings for Europe, following similar empirical strategies but unfortunately, almost never assessing the possible role of missing variables and reverse causality.

Bottazzi and Peri (2003) analyse the role of R&D and spatial technological spillovers on patenting in 86 European Regions for the period 1977-1995. The number of patents in a region is treated as a function

of R&D intensity in other regions located at various distances represented by classes: 0-300 km, 300-600km, 600-900km, 900-1300km and 1300-2000km. Importantly, R&D is instrumented using historical density in 1930. R&D spillovers are shown to diffuse within 300 km from the source region, while outside this distance no effects are found. Doubling the resources devoted to R&D in one region contributes to an increase in its own patenting activity of 80% to 90%. The same increase has a significant positive effect of 2-3% on patenting in regions within a 300km range and no effect any further away.

Similar conclusions are reached by Greunz (2003), who examines the distance effect in even more detail in a study of 153 European regions over the period 1989-1996. Inter-regional knowledge spillovers are supposed to decrease with the inverse of the square distance, and those emanating from direct neighbours, the neighbours of the neighbour (second order) and so on, are distinguished. All the local and the first three orders of neighbourhood significantly affect local patenting with decreasing intensity. Spillovers are essentially driven by the private business sector, while universities contribute mainly to local patents and first order neighbours. Importantly, technological proximity between regions is shown to matter too, and to reinforce spillovers.

Parent and Riou (2005) confirm these conclusions on a sample of 335 regions in nine European countries for the period 19891999, using transport time instead of distance. Bode (2004), in a study on Germany using spatial econometrics methods, shows that the contribution of an external R&D market potential is significant but low. Distinguishing regions with high and low R&D stocks, he also finds that only the latter benefit from spillovers. Finally, Maggioni et al. (2007), on a subset of 109 European regions, show that, on top of spatial spillover effects, being specialised in high-technology patents produces a higher number of patents per head. They also emphasise a significantly negative effect of being located at the periphery of Europe (measured by the distance from Brussels).

Proximity between patents and citations

Jaffe et al. (1993) and Jaffe and Trajtenberg (2002) follow another route, instead of estimating an innovation production function including spatial spillovers. It consists in comparing the location of patents and the location of patent use, as reflected in the citations they receive. The degree of spatial concentration of the latter is compared to a counterfactual that consists in the spatial concentration pattern of citations that would have emerged in the absence of technological spillovers. On US data, they find that domestic citations of patents are indeed higher than what would result from citations emanating randomly across space. This is shown to hold at the country, state and metropolitan area levels. This pattern decreases, but slowly, with time since patenting. For instance, at the metropolitan level, the overall citation matching of less than 10-year-old patents is 21.9 for top corporate firms, 8.8 when excluding self-citations, while random citations would produce 3.6 only. For a more than 10-year-old cohort of patents, figures are still at 13.3, 8.7 and 1.3 respectively. This kind of conclusion was confirmed for the US in a number of other studies, as for instance Thompson (2006).

Maurseth and Verspagen (2002) is a first study on 112 European regions for citations over the period 1979-1996. A gravity model of citations is estimated which leads to a number of interesting findings. Distance is proved to significantly reduce the flow of citations; citations within the same country, or between regions sharing the same language, also being higher even when distance is controlled for. As regards the role of industrial structure, citations also occur more often between regions that are specialised

in industrial sectors with specific technological linkages. Fischer et al. (2006), Fischer et al. (2009a), and LeSage et al. (2007), using Poisson models and spatial econometrics techniques, confirm these conclusions on the role of technological proximity. Verspagen and Schoenmakers (2004) also find that citations within multinational firms are more intense when establishments are closer in space, and this pattern also emerges between multinationals.

Maggioni et al. (2007) exhibit the role of distance in a gravity model of co-patenting over a set of 109 European regions. Technological distance is also included in the specification. The two notions of distance significantly affect co-patenting and there is a further negative effect of non-contiguity. Doubling the distance halves the co-patenting. Geographical peripherality (still in terms of distance from Brussels) does negatively influence the co-patenting activity. Increasing co-membership in a European research network increases co-patenting by 0.3%.

Finally, Fischer et al. (2009b) propose an exercise directly comparable to Jaffe et al. (1993) for Europe. They first illustrate the fact that there is a clear pattern of localisation for patenting at the regional and country levels. Then, comparing the location of patents and their citations, they show that citations are about seven times more likely to come from the same region as control patents, 2.6 times more likely excluding self-citations. They are also 2.7 times more likely to come from the same country as the originating patents, 1.7 times excluding self-citations.

2.2.5 Significance and explanatory power of structural models

The main conclusion to be drawn from structural estimation of economic geography models is that they are not rejected by the data. The Krugman-Helpman model is not rejected for United States counties (Hanson (2005)), for Germany (Brakman et al. (2004)) or for Italy (Mion (2004)), in the sense that the estimates are structurally consistent. They match theoretical constraints (for instance the elasticity of substitution between varieties has to be greater than one) and are close to the values generally admitted. It is also shown that the share of the spatial wage variance explained by this specification is fairly large and that the real market potential performs at least as well as, and often better than Harris' market potential.

There are certain limits to this approach, including, from the policy point of view, the fact that theoretical predictions of such a large scale model (with more than two regions) are not known. Moreover, the model only considers one differentiated good sector. Therefore the empirical application deals with wages aggregated over all industries, making it impossible to obtain industry-specific estimates or, consequently, industry-specific policy recommendations.

Given the assumption of absence of spatial labour mobility and the importance attached to access to intermediate inputs, this model is relevant to explain disparities at high spatial levels, typically between countries. Unfortunately, wage data are not available for a large set of countries or large regions. For this reason, Redding and Venables (2004) use GDP per capita as the dependent variable. They show that when the market potential variables are introduced into the specification alone, they both have a positive and significant effect on GDP per capita. They also explain more than 70% of the GDP per capita variance. Estimations are shown to be fairly sensitive to the choice of the internal distance used to assess the own-market size of the country. Typically, if the weight given to this market (corresponding to an ad hoc choice) is too high, one ends up explaining GDP per capita by GDP. This also demonstrates

again that both of the real market potential variables are probably endogenous. Redding and Venables (2004) consider two main strategies to tackle this issue. The first, more removed from theory, consists in removing the own-country from the market potential variable. The share of the variance explained by such real market potential external to the country is still equal to 35%. This also makes Canada richer than the US, for instance, which is not very appealing in terms of either realism or policy. The second strategy uses the distances to New York, Brussels and Tokyo, which - maybe surprisingly given their obvious correlation with current productivity shocks - pass over-identification tests. Like Hanson (2005), similar results are obtained when many control variables are introduced[1]. The main problem that Redding and Venables (2004) face is that, unfortunately, the real market potential variables are never simultaneously significant, which would correspond to the structural model. This is probably due to the excessive correlation of the two market potential variables. Strictly speaking, it prevents us from deciding whether or not to reject the model using structural parameters, which cannot be recovered. After the first application to disparities between countries at the world level by Redding and Venables (2004), this strategy (but without the role of intermediate inputs) has been applied to various regions to explain European regional GDP per capita (Breinlich (2006)) and wages (Head and Mayer (2006)). Lastly, like Hanson (2005), the study is performed on aggregate data, making it impossible to provide industry-specific policy implications.[2]

[1]Redding and Venables (2004) deal with primary resources, geography, and institutions but not education as Hanson (2005) does

[2]See Combes, Mayer and Thisse (2008) for more details.

Chapter 3

Direct assessments of regional policies (empirics)

As we will see in the last chapter of this report, combining theoretical results from economic geography models and the empirical assessment of the magnitude of agglomeration effects allows us to derive some implications for regional policy. An academic literature chooses another route that consists in directly assessing the impact of some regional policies that have been implemented in various contexts. We detail in this chapter the role that is found for public infrastructure, tax differentials, European regional funds, and for some national grants for regional assistance. Before, we briefly recall the challenges that are met when one seeks to evaluate directly the impact of a policy.

3.1 Generic empirical concerns and possible solutions

Assessing the impact of a decision, for instance using a medicine, on the outcome that it is supposed to influence is difficult in any field. Economic contexts, due to all the market inter-dependencies that characterise them, make it in general even more tricky. A large literature has been designed to provide a correct approach to this question but detailing it goes well beyond the scope of the report. The interested readers can find for instance in Heckman et al. (1999) a pretty complete overview of the problems that policy evaluation in economic situations raises. In this section, the purpose is only to illustrate the issues at stake and to provide the intuition of the solutions that can be proposed, before presenting in the following sections those that have been implemented for the case of regional policies.

The problem is the following. One decides to treat part of the population, in our case to provide a grant to a region, or to some firms in a region for instance, and one wants to assess if this treatment has an effect or not. The job would be easy if one could compare the outcome of the region once treated, which one can in general observe, to the outcome of this very same region if it would have not been treated. However, and by definition, the latter outcome is not observable since the region has been treated. A natural strategy is to compare the outcome of treated regions to the outcome of other non-treated regions. The big issue is to determine when this comparison is relevant or not.

If ex-ante all possibly treated regions are identical and the treatment is randomly applied to some of them only, the comparison is relevant. There are no reasons for which the treatment would have had a different effect on the treated and non-treated regions, and therefore a non-treated region is identical to a

treated region that would not have been treated. The problem is that is general the treatment is applied to some sub-populations that are different from the sub-populations that are non-treated. For instance, to assess the impact of a new drug, one cannot compare the health status of the hill patients who have been treated to the health status of non-hill patients that were not treated. Clearly, the latter are in general in better shape, precisely because they were not treated because in first place they were not hill. One has to compare "identical hill patients" that have been and have not been treated. Recognising this leads to a first solution to the problem, which is the most used to assess the impact of new drugs. It consists in treating only a share of the hill patients, chosen randomly over all hill patients, and to compare them to the non-treated patients. Because simply knowing that one is treated may have an impact of the treatment efficiency, the patients must not even know whether they are treated or not. Therefore a placebo, ie a treatment that looks like the real one but has no effect, is delivered to the non-treated.

When one comes to economics, the fundamental problem is the same since, in general, the treatment, a fund provided to a region, is applied to regions with characteristics that are different from the non-treated regions. For instance, regional aid is reserved to poor regions, where GDP per capita is below a certain threshold. If the purpose is to increase GDP, comparing the regional GDP of the regions who benefit from the funds to the GDP of other ones is little delivering. The latter are probably still richer since precisely the funds were given to the poor regions. The effect of a dummy variable "treated" in a regression that explains GDP per capita would thus be negative, not meaning that the treatment had a negative impact but only that the regions did not fully catch-up. If it is positive, one can infer that the treatment had a positive impact (the regions are now ahead while they were initially behind) but the magnitude of the policy impact is not known again. They can even be ahead just because a very positive shock other than the policy hit only less rich regions. In any case, the coefficient tells nothing about their backwardness if the fund would have not been provided.

Following medicine practice, one could chose to provide funds to some of the poor regions only and then to compare the ex-post GDPs of the poor regions who benefit from funds to the GDP of those who did not. However, first, this rises some equity questions, as in the case of drug administration, since some poor regions will not have access to the funds (as some hill patients do not have access to a treatment that could have improved their health status). In medecine and when sample size are small, one accepts this constraint in general, but it would be probably much more difficult regarding regional policy in Europe for political reasons. Second, knowing that the region is not treated may have an impact on its GDP, be it because agents will be even more depressed and lose even more GDP in this region, or, on the contrary, because the agents anticipating that they can count on them only will make more efforts (and conversely in treated regions, they would make less efforts), which will in turn increase (decrease respectively) their GDP. Unfortunately the placebo solution is not available for economic policy. Still, such a random experiment solution is sometimes implemented for economic policy but more at the firm or the household/village level. Some agents benefits from the policy while others do not, and one compare their ex-post outcomes, hoping that being or not treated has no indirect effect other than the one one wants to assess. Esther Duflo has a number of contributions (see Duflo et al. (2007) for a survey) that advocate a lot such experiments to assess the impact of development policies in poor countries.

Using econometrics enlarges the comparison possibilities between treated and non-treated agents in order to make comparisons more relevant. Instead of directly comparing them, one can regress the targeted economic outcome not only on the fact to be treated, or the magnitude of the treatment, but also on

observable variables other than the treatment that can affect it. Typically, instead of comparing the GDP of the regions who benefit and do not benefit from regional funds, one simply regress the ex-post GDP not only on a dummy taking value 1 if the region got some funds, zero if not, but also on the initial GDP. Now the assumption is "only" that the outcome was not influenced by any variables other than the initial GDP and the treatment. If this is satisfied, the effect of the dummy corresponds to the true effect of the policy. However the initial GDP is not enough to control for regional differences that can be correlated with both the fact to have access to regional funds and the ex-post GDP. For instance the industrial composition, the skills level of the workforce have clearly also an impact on it. Therefore, one has to control for all of such characteristics. The problem is that in general some of them are not observable by the econometrician. Imagine for instance that the low level of GDP in a region is due to a bad economic governance of local politicians but that these politicians are efficient in getting the grant (the time they do not spend to monitor the economy is spent to lobby for getting the grant). The fact that politicians badly govern enters the residual of the specification because it is not observed by the econometrician. Since it is correlated to the fact of getting the funds, the ols estimate of the fund impact is biased. In this precise example, one can even tell that it is underestimated since it captures the negative quality of the politicians of the regions who get the grant.

Economists propose a further solution called the difference-in-difference estimator. When observations are repeated over time, we can not only compare the outcome of treated and non-treated regions but also the time-variations of their outcome. It is exactly identical as introducing a region fixed-effect in the specification. This fixed-effect captures the role of any variable specific to the region (observable or not) constant over time. For instance, it can capture the quality of the local politicians. But it still assumes that this quality does not change over time, more generally that the regional shocks at each date are not correlated with the policy. If it does, and again this may result from the implementation of the policy it self, the difference-in-difference estimate is still biased.

Instead of controlling for observable variables in the specification, one can attempt to match treated and non-treated agents according to some characteristics they share such that keeping in the sample only "comparable" observations. This strategy can be based on the use of a couple of variables but a more efficient strategy has been proposed. The idea consists in computing what is called a propensity score for each agent or region, which corresponds to the probability to be treated, conditionally on some characteristics. One first estimates a probit model of this probability regressing the fact to be treated or not on all characteristics that can affect this probability but do not affect the outcome to be influenced by the policy. One can then recompute for each observation the probability to be treated according to the model, independently to the fact that one is actually treated or not. Finally, one keeps in the sample on which the impact of the policy is evaluated only the observations that correspond to intermediate probabilities of being treated. This assumes that when the probability to be treated is either too high or too low, the agent or region has some characteristics that too much differ from the characteristics of those who are / are not treated. Conversely, those kept in the sample, treated or not, should be relatively similar. Note that such a strategy does not prevent the researchers from using a difference-in-difference estimator. Indeed the difference-in-difference estimator still assumes that the error term in the specification is drawn from the same distributions for observations corresponding to the treated and the non-treated agents. Matching them ex-ante using a propensity score allows keeping only those that are not too different along some characteristics, which helps to have this assumption satisfied.

In the same spirit, some authors propose to appeal to a solution named "regression discontinuity". The idea consists in keeping only those observations that are close to the threshold of the criteria that makes that one is treated or not. For instance for European regional funds, and as Becker et al. (2010) proposes, one can keep in the data set the regions that are only slightly above or below the threshold of eligibility which is at 75% of the European average regional GDP. Doing so, one selects regions that are probably similar in many other respects, which could be observable or not, and that fell to one or the other side of the threshold mostly randomly. In this case, comparing the ex-post situation of treated and non-treated regions is relevant since those are comparable non only as regards possibly observable variables but also for some unobservable variables. The issue here is that one is never sure that the conclusion reached about the validity of the policy, which is obtained in the neighbourhood of the threshold, would be also valid for observations that are not in this neighbourhood.

Finally, since the concern is the presence of a correlation between the fact to be treated and the residual of the specification that evaluate the determinants of the outcome targeted, which biases the OLS estimate and is mainly due to reverse causality and missing variables issues, one can use the standard tool proposed to deal with such biases, instrumentation. The difficulty is, as always, to find relevant instruments. They have to both explain well the fact to be treated or not but not the outcome of the policy. In other words, one has to find instruments correlated with the regions characteristics that do neither enter the outcome specification nor influence its unobservable determinants but influence the fact to be treated. We will present below a number of studies choosing this strategy. Again, many methods can be combined to try gaining on all sides. For instance, Criscuolo et al. (2012) propose an instrumental variable method of a difference-in-difference model with a matching ex-ante selection of observations removing outliers.

Note that the strategy to control for the possible biases that arise from applying the policy not randomly across observations does not deal with the possible endogeneity of other variables present in the specification. Typically, the studies that tackle the possible endogeneity of the policy instrument never simultaneously tackle the possible endogeneity of control variables, as for instance the overall size or density of the region or its industrial composition, even if chapter 2 in this report insists on all the reasons why it should be done.

3.2 Public infrastructure

A first set of possible regional policy instruments that has been investigated in the literature regards the possible role of public infrastructure. As before, studies at the European level are rare. Basile et al. (2008) and Basile et al. (2009), who consider Nuts 2 regions of eight and five European countries respectively, study the impact of an infrastructure index that takes into account the regional roads, railways and telecommunication infrastructure. The former find that it has a negative impact on the location choices of European multinational firms and no significant effect on non-European multinationals, while it has non significant impacts on both according to the latter. Spies (2010) also find no significant effect on the location choices of FDI in Germany of a local infrastructure index constructed out of the relative length of highways, roads, rivers and the number of airway passengers. The same for the regional road endowment is obtained by Wren and Jones (2011) for the FDI location choices in the UK. Pusterla and Resmini (2007) working on Nuts 2 regions of four eastern countries, Bulgaria,Hungary, Poland and Romania even find a significantly negative impact of public roads per capita on FDI location choices in these countries.

A couple of other studies are a bit more favourable to the role of transport infrastructure. Hilber and Voicu (2010) find that FDI location choices in Romania are positively influenced by railroad density. However the effect is not robust to the introduction of regional fixed-effects, which would either mean that there is not enough variability to identify such an effect on time variations only, or that the effect of infrastructure is associated to the presence of some of the permanent characteristics of the regions. Basile (2004) finds for Italy a positive effect of an infrastructure index of roads, railways and telecommunications on both acquisitions and greenfield investments. Barrios et al. (2006) for Ireland also obtain a positive impact on FDI of the closeness to airports and ports in the seventies, but no significant impact nowadays. Lastly, Holl (2004) for Spain and using micro-level data and geographic information system techniques finds that new motorways do affect the location of new manufacturing establishments (foreign owned or not) at the municipality level.

Taken at face value, there does not seem to be any large effect of local infrastructure on firms location choices. However, first notice that considering infrastructure as an input of the local firms production function as these studies implicitly do, is very much in the spirit of the first endogenous growth studies but it ignores one of the most important message of economic geography. Transport infrastructure is not only a local input but it is also a mean to connect different regions, facilitating trade between them. Improving infrastructure in a region can clearly be simultaneously beneficial, or detrimental, to other regions. Martin and Rogers (1995) emphasise that, if improving local infrastructure mainly benefits to the region where it is improved, improving inter-regional infrastructure can be beneficial for one region only, the larger or the smaller region depending on which side of the bell-shaped curve the economy is. From the empirical point of view, this implies to distinguish intra- and inter-regional infrastructure in the specification, which is never done. Importantly, it also requires to interact the role of infrastructure with the distribution of activities, as embodied in a market potential function for instance. Again, it is not done. Finally, Duranton and Turner (2011) show that instrumenting the infrastructure stock is really crucial to get relevant estimates of its impact on US cities growth. Once this carefully done, they obtain that a 10% increase in a city's initial stock of highways causes about a 1.1% increase in its employment. Clearly, roads are build either in expanding regions that suffer from congestion or, on the contrary, in depressed / peripheral regions (with the prior that it would help them to catch-up): both induce a possible reverse causality bias. However, none of the previous studies on European countries explore this issue (except the introduction of regional fixed-effects in Hilber and Voicu (2010) which is a first step in this direction).

3.3 Tax differentials

There are few studies that study the role of local taxes on regional outcome in Europe. For the US, the pioneering study of Head et al. (1999) found that lower corporate taxes do significantly attract Japanese investments, at the state level. Their simulations indicate that unilateral withdrawal of promotion would have caused individual states to lose substantial amounts of Japanese investment. However they indicate also, because state promotional policies tended to offset each other, that their impact on the geographic distribution of Japanese investment is small. As regards European regions, Head and Mayer (2004) introduce the role of both social charges rate and of corporate tax rate separately to explain Japanese FDI location choices. They both have a significantly negative impact on location choices but it is not

robust to the introduction of country dummies. A nested model shows that both, once averaged at the country level, do negatively impact the choice of the country by the firms. Basile et al. (2008) and Basile et al. (2009) also estimate a significantly negative effect of corporate tax rate for investment over European Nuts 2 regions for both European and non-European multinationals. The tax wedge on employment has a positive effect on European multinationals in Basile et al. (2009) while it is non-significant in Basile et al. (2008) and it has a negative for non-European multinationals in both papers. Spies (2010) who studies FDI in Germany finds no significant effect of either the business tax and the real estate tax rates.

Again, the possible endogeneity of the tax is not assessed in the previous studies. However, if one has in mind the tax competition models we present in chapter 4, clearly, first location in a region depends not only on the tax level in this region but also in all other regions through the impact that these taxes have on all other firms location choices. Second, the tax level chosen in a region depends on the characteristics of these regions be it only through the regions budget constraint, and therefore on the location choice of firms, which induces a possible reverse causality bias.

Two articles, Rathelot and Sillard (2008) and Duranton et al. (2011), consider such possible biases seriously. They study the role of local taxation on firm creation in France and the UK respectively. They appeal to the same strategy that consists in restricting the sample to pairs of firms located very close to each other (less than 2 kilometres) but in different regions, ie firms that are closer than one kilometer from a regional boundary (even if the pairs of close by firms located in the same region can be kept to improve the efficiency of the estimate of the impact of variables other than tax rates). The regional level is Nuts 3 for France ("départements"), while this is the "jurisdiction" level for the UK. As a result, the two firms constituting each pair are such that they face different tax rates but similar economic conditions in terms of market access, characteristics of the labour force, ie a number of variables that can be observed and introduced in the specification, but they should be also similar in terms of unobservable variables. This borrows both to matching methods and regression discontinuity since the strategy exploits the tax rate spatial discontinuity due to the presence of administrative borders and that firms are matched by pairs. The strategy is further improved. First, both Rathelot and Sillard (2008) and Duranton et al. (2011) instrument tax rates to consider their possible correlation with the part of the random component not removed by matching. The instrumentation strategy is more convincing in Duranton et al. (2011) who appeal to local political variables (typically the share of local politicians affiliated with the three main political parties, a set of dummies indicating whether the local authority is controlled by one of the three main parties and a set of interactions giving the share of the three main parties if they control the local authority). Rathelot and Sillard (2008) simply use the sum of departmental and regional tax rates differentials.

Actually, the main contribution of Duranton et al. (2011) consists in the fact that the strategy is applied not only to firms creation but also to employment growth. Employment growth at the firm level present the further advantage to simultaneously appeal to the instrumentation and the advantage of the difference-in-difference approach.

Rathelot and Sillard (2008) find that higher local taxes reduce firms creation but the effect is weak, such that increasing the tax rate differential by one percentage point increases the probability of a firm setting up in the lower taxed municipality by around one percentage point. Duranton et al. (2011) find that local taxation of non-residential property has a sizeable negative impact on employment growth, but no effect on entry. Importantly they show that methodologies that do not address the three problems

of individual heterogeneity, unobserved time-varying location-specific effects and endogeneity of local taxation, give substantively different results.

3.4 European regional funds

The literature studies the impact of European structural funds on two distinct variables: FDI on the one hand, regional growth on the other hand. The role of two types of variables is investigated: the impact of the overall amount of structural funds received by the region and the impact of being eligible to these funds, possibly distinguishing the type of eligibility in terms of Objectives 1, 2, 3,... As for other regional policies, results found are mixed in general and endogeneity is rarely seriously tackled.

In their study on Japanese FDI across European regions, Head and Mayer (2004) do not find any impact of the objective 1 eligibility on the firms location choices. This is confirmed by Basile et al. (2009) and Basile et al. (2008) who do not find any significant impact of objective 1 eligibility on the location choices of both European and non-European multinationals. Structural funds do not have either a significant effect on European multinationals location choices but they significantly impact the location choices of non-European multinationals. The reverse is found for being a recipient country of the Cohesion Fund, which has a positive impact for European multinationals only. Crozet et al. (2004) on France are able to distinguish the impact of the various objectives on the FDI location choices. They find no impact of objective 2 grants, objective 5b grants, and of community initiative grants that the European Commission also provides.

Quite a few papers investigate the extent of the influence of the European funds on the regional growth and convergence process. The sample and time period differ but the main conclusion remains: no strong effect in either direction is found. On the 1989-1999 period and a sample of 145 Nuts 2 regions, Dall'erba (2005) exhibits a positive influence of structural funds on regional growth. This is partly confirmed on a sub-sample of 41 regions by Puigcerver-Penalver (2007), who estimate a significant effect of objective 1 eligibility on growth for sub-period 1989-1993 but not for 1993-1999. Rodríguez-Pose and Fratesi (2004) found no general effect of objective 1 structural funds on growth, only the funds related to education and human capital investments, which only represents one-eight of total commitments, having a significantly positive effect. Ramajo et al. (2008) perform regressions for regions belonging to cohesion countries and non-cohesion countries separately. They find that convergence is faster between the former. Finally, Esposti and Bussoletti (2008) obtain slightly more optimistic results with some weak evidence of a positive impact of objective 1 funds on growth. It however depends on the country where regions are located, the largest effect being obtained for French regions. Non-significant and even negative effects are estimated for others.

None of the previous studies on the impact of European structural funds and eligibility considers the possible endogeneity of the funds with respect to FDI or growth. As we repeatedly pointed times in this report, such estimations are possibly plagued by both missing variables and reverse causality. Two exceptions are worth mentioning. Dall'erba and Le Gallo (2008) appeal to a spatial econometrics estimation and both internal and external instruments. Internal instruments refer to, as is standard in this literature, some spatial lags of all explanatory variables. External instruments are the distance by road to Brussels, the travel time from the most populated town of each region to Brussels and some dummy variables (and their spatial lags) for the position of the region in the distribution of some of the

explanatory variables. Importantly, not all explanatory variables are instrumented (but regional funds are) and, while over-identification tests are available for these technics and passed here, weak instruments tests are not. Beyond these limits, it is found that regional funds do not have any influence on the convergence process of European regions. Moreover, simulations show that investments targeted to the peripheral regions do not spill-over regional boundaries. Appealing to a time panel data set, Mohl and Hagen (2010) propose to use the dynamic panel econometric method (GMM) to deal with the possible endogeneity of European funds. They find that if the total amount of funds do not have any significant impact on regional growth, objective 1 do promote regional growth. Similarly to the previous study, the GMM approach suffers from its kind of black box perspective. No real economic intuitions sustain the validity of the instruments and no test assesses if they are not weak.

To address this concern via another route, Becker et al. (2010) design an interesting regression discontinuity setting based on the fact that eligibility to objective 1 funds is constrained by being below a GDP threshold equal to 75% of the European average. The authors select a sample of Nuts 2 regions that are just below and above this threshold. They complete this strategy by performing their estimation at the Nuts 3 level. This increases both the variability of the outcome and of the probability to benefit from the policy, and the comparability of the treated and non-treated samples. Indeed, some Nuts 3 regions within the treated Nuts 2 regions (thus below the 75% threshold) may well be themselves above this threshold, and therefore richer than other Nuts 3 regions present in the sample with a GDP below the threshold but not eligible because belonging to Nuts 2 regions above the threshold. Finally, because a time dimension is available in the data set, a difference-in-difference estimate is computed, which controls for regional fixed effects. It is found that objective 1 funds do have an impact on regional growth. The programme participation exerts a differential impact on GDP per capita growth of about 1.8 percentage points within the programming period. With respect to employment, a significant, but smaller, positive effect of about 0.5 percentage points is found. The authors propose a back-of-the envelope calculation that suggests that the funds spent on Objective 1 have a return which is about 20% higher than their costs. These conclusions are interesting but remain to be confirmed since, unfortunately, Becker et al. (2010) do not instrument the policy itself, even if this could be less necessary than in Dall'erba and Le Gallo (2008) and Mohl and Hagen (2010) due to the difference-in-difference regression discontinuity strategy. Moreover, the main limit of regression discontinuity designs remains, that, strictly speaking, results are valid only for those observations in the neighbourhood of the discontinuity.

3.5 National funds for regional assistance

A last series of papers attempt to assess the effect of some national regional assistance programmes. For instance, Barrios et al. (2006) study the role for FDI of the public policies in favour of the Irish "designated areas", which correspond to underdeveloped territories. They find that this policy has been an effective tool in promoting the location of foreign plants in these counties for low technology firms and since the mid-eighties only. The programme dates back to the fifties. Moreover, for high technology firms, agglomeration economies are a more important location determinant than public incentives. Never the less, no attempt is done to tackle any of the, already mentioned, possible biases that can affect the evaluation of the policy role. This is a severe limitation for the credibility of the results.

Wren and Jones (2009) compare the survival duration between greenfield investments and re-investments in the UK. Survival rates are apparently longer for the latter but the authors show that it only results from a selection of more efficient plants in this kind of investment. No difference in survival rates emerges when selection is taken into account. As regards regional grants, the role of which is also estimated conditionally on selection, they appear to have no impact on survival rates either, even if they lead to more re-investment.

Martin et al. (2011a) is a first study more achieved in terms of consideration of the various possible biases affecting the evaluation of national programmes. The authors analyse the role of local productive systems in France. This is a regional policy tool that provides grants and tax exemption to firms located in some designed areas with the rational of stimulating the formation of "industrial clusters" where agglomeration effects are magnified, as advocated by various contributions of Porter for instance (Porter (1998), Porter (2000)).

Martin et al. (2011a) first show that the policy clearly targets firms located in backward regions and operating in declining industries even if the official objective was to promote agglomeration externalities and to re-orientate the French regional policy from less equity concerns towards more efficiency objectives. From the policy evaluation point of view, this underlines the importance to control for such a non-random selection of the firms entering the programme. The impact of the policy is studied on TFP, employment, exports, and survival rates. The strategy mainly consists in a difference-in-difference approach comparing the time variation (before and after the policy was implemented) of TFP (or employment or exports) of firms who benefited and did not benefited from the policy. In other words any firm specific fixed-effect, which includes the non-time varying characteristics of its location, is removed. A first extension consists in allowing for the presence of an auto-regressive effect in the error term. Alternatively, a triple-difference strategy is also proposed. It compares the growth rates of these firms before and after the implementation of the policy, which removes any firm-specific time trend. Finally, all these strategies are completed by a matching approach that reduces the sample to those firms sharing comparable propensity scores in terms of their probability to enter the programme.

Conclusions about the efficiency of the programme are pretty pessimistic. Martin et al. (2011a) find no significant effect of French local productive systems on TFP, employment and exports, nor it does affect the survival rate of firms.

Mayer et al. (2011) study the impact of a French Enterprise zones program on establishment location decisions over the period 1995-2007 at a very small geographical scale, the census block level. The programme, called Zones Franches Urbaines (ZFU), aims at the revitalisation of French urban fringes by giving massive tax breaks for firms locating in those are. Mayer et al. (2011) adopt a difference in difference approach that combines spatial and time differentiation. Alternately they also use the fact that targeted urban areas have been selected in different waves over time and they also exploit a discontinuity in the eligibility criteria of the policy as an exogenous source of variation to estimate the impact of the assistance programme. However, they do not instrument the policy itself.

The policy does have an impact on the probability that establishments locate in targeted areas. Importantly, they show that the impact of the policy is stronger for targeted areas that are initially less distressed and for sectors in which relocation costs are lower. Moreover, ZFU areas tend to attract smaller firms. However, the analysis of the spatial pattern of the effect reveals that the policy does not create economic activity per se but rather operates as a firms relocation device within municipalities, inducing

existing establishments, or new establishments to (re)locate in the ZFU part of the municipality.

These results are broadly consistent with the findings of Briant et al. (2011) for the same French assistance programme, who use a non instrumented difference in difference strategy. They find that only a small positive average impact on firms and jobs creation rates that is strongly heterogeneous across targeted neighbourhoods. They show that the geographical characteristics of the neighbourhoods account for part of this heterogeneity. Spatial isolation, which account for urban severance and transport access, makes the programme less efficient.

Even if Martin et al. (2011a), Mayer et al. (2011), and Briant et al. (2011) constitute a significant progress in the consideration of the possible biases related to regional policy evaluation, one can still argue that some endogeneity of the policy itself may remain. As the authors themselves acknowledge, instrumenting the difference-in-difference specification, as it is done for taxes by Rathelot and Sillard (2008) and Duranton et al. (2011) for instance, would even more lead credence to the results obtain. It is however difficult to find good instruments but two studies, Devereux et al. (2007) and Criscuolo et al. (2012), manage to do it for the evaluation of the regional selective assistance programme implemented in the UK.

Devereux et al. (2007) estimate the impact of regional assistance on the location choices of greenfield investments by both domestic and foreign firms over 88 UK counties, some being divided in their assisted and non assisted parts. Contrary to most studies, the policy variable that is studied is not the eligibility of the area where the firm chooses to locate its investment but the amount of the grant received by each firm, at the individual level so.

Clearly, if one gains in terms of the precision of the policy evaluated, possible endogeneity concerns are reinforced since the grant received can clearly be correlated to some unobserved characteristics of the firm that also impact its location choice. The strategy consist in noticing that, in logit models, the firms location choices do not depend on the own firm time invariant characteristics since such characteristics, would they be observed or not, take the same values for all possible locations. By contrast, these characteristics, as firm size, industry or the efficiency of the managers, clearly affect the fact to get the grant and the amount obtained. Therefore they can be used as instruments to predict the grant, on top of some, or even all, characteristics of the area or area-industry that also influence the location choices. Therefore Devereux et al. (2007) first estimate a model relating the grant received to the firm individual characteristics as well as to some of the standard agglomeration effects we described above. They use a two-step tobit model since the amount of the grant is predicted conditionally on getting a grant. The exclusion condition of this model is based on variables that are supposed to reflect the cost to apply (firm size for instance) but not the magnitude of the grant received (that depends on the size of the project, not of the firm).

This first step strategy allows Devereux et al. (2007) to obtain a predictor of the grant that (i) is orthogonal to the individual firms shocks that affect their location choices (ii) is available not only for the location for which the firm got the grant eventually but also for all possible locations. Then the logit model is estimated using the predictor of the grant as an explanatory variable next to all agglomeration effects.

It is crucial to notice that this estimation strategy is valid as long as the firm characteristics do not interact with the local characteristics to explain location choices. For instance, if the firm size interact with the size of the economy to benefit from agglomeration effects, then the grant predictor may remain

correlated with the random element in the logit model. This is probably a pretty strong assumption, since, for instance, Combes et al. (forthcoming) show that more productive and larger firms do benefit more from agglomeration effects.

The estimates obtained by Devereux et al. (2007) implies that grants do have a significant effect on average in attracting plants to specific geographic areas. However, the marginal effect is very low, implying that an increase in the expected grant of £100,000 is associated with a 1% increase in the probability of location, which evolves for instance from 1% to 1.01%. Importantly, firms are found to be less responsive to government subsidies in areas where agglomeration effects from own industry are weak. Including the interaction term between the grant and the local number of plants in the industry, the average marginal effect of the expected grant is estimated three times higher. But the interaction term indicates that as the local size of the industry rises, the marginal effect of the expected grant does also. An increase of 10 plants increases the marginal effect of an increase of £100,000 in the expected grant on the probability of location by 6.7%. Therefore higher grant offers are needed to attract greenfield entrants to locations where industry agglomeration or natural resource benefits are weaker. This suggests that subsidies are less effective in influencing firms' location decisions in the face of countervailing co-location benefits.

One limit faced by both Devereux et al. (2007) and Martin et al. (2011a) relies in the difficulty to match the data set reporting the firms who benefited from the policy and the data sets that report their individual characteristics. At the end, the sample size on which the identification is based is really small (a couple of hundreds of firms) even if estimation is performed on much larger numbers of observations, which however only improves the efficiency of the estimates of the variables other than the policy one. Moreover, Martin et al. (2011a) do not control for time-varying agglomeration effects (time-invariant ones are controlled for since they enter the region fixed-effect that disappears by differencing). Devereux et al. (2007) do control for them but assume their exogeneity.

To address both concerns, Wren and Jones (2011) use data aggregated by location and instrument all explanatory variables. There is cost to that approach as a lot is lost in terms of the precision of the information used and the individual controls considered, even more in Wren and Jones (2011) since these authors use a very aggregated spatial classification considering only 10 regions for the UK. Once these limits acknowledged, when a time panel is available, one can appeal to GMM estimation technics to instruments. We have underlined above the possible limits from which they can suffer, in particular as regards possible weak instruments concerns.

Wren and Jones (2011) explain the location choices in the UK of new projects that include start-up, acquisition, joint ventures, and reinvestments. Instrumenting by GMM, they estimate a significantly positive and concave effect of the RSA grants. Each £25 million of grant changes the regional location of about six inward FDI projects. On average, projects have 150 jobs and each job diverted costs £27,500. Among other local characteristics variables that impact location choices, past FDI is found to have a strong positive effect. The impact of other variables is less intuitive, as the negative impact of population and the positive ones of unskilled workers and of wages, or the absence of effect of the distance to major cities and of the skilled population.

Criscuolo et al. (2012) also evaluate the impact of the regional assistance programme in the UK but on employment and investment. As Devereux et al. (2007) they work at the individual level. On the top of that they exploit the time dimension of their data set. They mix a difference-in-difference approach completed by matching with a simultaneous instrumentation of the policy implemented. Contrary to

location choices in logit models, individual firm effects can affect both the grant obtained and the firm employment or investment. Therefore, the Devereux et al. (2007) instrumentation strategy is not valid anymore. To instrument the policy, Criscuolo et al. (2012) appeal to the eligibility of the locations where the firm has some plants and use the fact that, due to constraints from the European Commission that evolve over time, this eligibility also changes over time. Many dummy variables corresponding to different levels of maximum percentage of investment eligible (which differs across areas and time) are used as instruments once averaged at the firm level over all the firm's establishments. These instruments are clearly correlated with the likelihood of a firm to receive a grant but not with the individual firm performance, would it be observed or not. Some concerns regarding the endogeneity of the instruments would only rise if, for instance, eligibility disappears precisely because employment or investment conditions improved. To reduce such a concerns, instruments are also lagged, typically by two years. To test for the validity of the instrumentation strategy, instruments are further completed by allowing for plant-specific trends and theoretical instruments based only on policy-changes rules. The difference-in-difference part of the Criscuolo et al. (2012) strategy is close to the one we described for Martin et al. (2011a). It reduces to specify the firm employment or investment at each date as a function of both the policy implemented at that date and a firm fixed-effect. Here again a propensity score for the probability to be treated is used to trim ex-ante extreme observations from the sample and reduce it to firms that should be more comparable.

Criscuolo et al. (2012) find that there is a large and significant average effect of the UK regional assistance programme for employment, investment and the probability of exit. These effects are seriously underestimated if endogeneity is ignored, as the participants in the programme appear to be weaker firms who would otherwise perform badly given their observable and unobservable characteristics. Importantly, there appear to be conversely no additional effects on productivity after controlling for the investment effects. Since the proportion of employment in entrants as a whole is shown to fall in areas benefiting from the policy, this raises the possibility of negative aggregate productivity effects from lower reallocation, resulting from the programme implementation, to more efficient firms due to the protection of inefficient incumbents. Another important result is that the program has an effect on firms that is differentiated according to their size. Only small firms benefit from the program. On the other hand, they estimate that the cost per job of the program is only $6,300 suggesting that investment subsidies can be cost effective.

Finally, note that Accetturo and de Blasio (2012) perform an aggregated version of this strategy to assess the impact of another regional assistance policy implemented in Italy. Only employment and firms creation effects are studied and data is aggregated at the municipality level from the beginning. They implement a two stage strategy: a matching procedure removing outlying municipalities followed by a difference-in-difference procedure that removes municipality fixed-effects. Finally, the policy is also instrumented by eligibility criteria at the European level. Aggregation makes a priori endogeneity concerns induced by time-varying local shocks possibly correlated with the policy more serious than in Criscuolo et al. (2012). The results suggest that the programme has been largely ineffective, as the growth in employment and firm numbers in the municipalities involved in the policy does not differ significantly from that in the cities not involved.

Chapter 4

Regional aids, competition policy and fiscal competition (theory)

4.1 Fiscal competition and fiscal externalities

The widespread phenomenon of good and factor market integration makes countries compete for most of their tax bases. For instance, when firms or capital are mobile across countries, it is likely that their location is influenced by each country's corporate tax. One may be interested in the extent to which such competition is desirable. We shall not address here the more consensual question about the efficiency-enhancing role of production factor mobility in a distortion-free world. We shall instead focus on the desirability of letting countries decide freely their fiscal policies, as opposed to some form of harmonisation or centralisation of fiscal decisions.

Let us begin by defining the scope of our survey. Mostly, the literature has focused on tax competition, i.e., settings in which the fiscal policy set by jurisdictions is the tax rate. Other policies have been studied, as, for instance, inter-regional transfers Hindriks and Myles (2003), public good level Wildasin (1988), Wildasin (1991), public inputs that improve the quality of capital Keen and Marchand (1997) or enforcement of income tax together with income tax level Cremer and Gahvari (2000). We use the term *fiscal competition* to encompass these different instruments. As regards tax competition, both direct and indirect taxation are studied in the literature. The former rests on the mobility of production factors, and the latter on cross-border shopping. Factor mobility may focus on (skilled or unskilled) labour or capital. Fiscal competition may arise among governments of any spatial entity that has taxing powers over a mobile tax base (e.g., countries competing for capital or cities competing for residents). We shall focus on countries competing for capital and/or firms, but some of the insights below apply to other fiscal competition contexts. Finally, one may distinguish between horizontal and vertical fiscal competition, the latter referring to an immobile tax base which is taxed by different layers of government (e.g., the local, state, and federal one). In what follows, we shall concentrate on horizontal interactions between countries. Internationally mobile production factors may be taxed according to the residence (i.e., where the production factor owner resides) or source principle (i.e., where the production factor generates income). While the residence principle would eliminate (most of the) distortions described below, it is generally recognised that corporate taxation is *de facto* source based Keen (1993). The survey thus focuses on source-based capital taxes.

One may distinguish two strands of the literature, which imply different welfare implications of fiscal competition. We shall focus here on the overall welfare, that is, that of the group of countries. Naturally, an individual country may gain from tax competition, an issue which has been raised in the literature and which may be one of the reasons behind the lack of harmonisation policies observed in reality (see, e.g. Peralta and van Ypersele (2006)).

Tiebout tradition The so-called Tiebout hypothesis Tiebout (1956) states that if individuals can move freely across regions, then they "vote with their feet", i.e., each chooses to live in the jurisdiction that provides him with the local public goods and corresponding taxes that best suit his preferences. Competition among jurisdictions ensures that the public goods are optimally provided. Although the original Tiebout argument focused on individuals, it may also be applied to firms. Competing governments providing efficiency-enhancing public services (e.g., infrastructure) to footlose firms end up charging the firms with the marginal cost of the public service, thus ensuring efficient provision and efficient sorting of firms across jurisdictions. The Tiebout argument is simple and powerful, because it rests on an analogy between competing firms and competing governments. We know, however, that the conditions for it to apply are very stringent. In particular, it precludes any type of strategic behaviour on behalf of governments (as with profit maximising firms operating in a perfectly competitive market). In other words, the governments should take the equilibrium level of profits as given, and not foresee their policy decisions as influencing the firms' profits.

Fiscal externality When each government reckons that the firms' profit level, or the return to capital, is affected by its policy choices, we move away from the Tiebout world. When a given country's policy choice changes, e.g., the world net-of-tax capital price, it affects the tax base of the other countries – the so-called *fiscal externality*. As a consequence, different sorts of inefficiencies are bound to arise. We review them below.

4.1.1 The race to the bottom

The basic result from this literature is that capital tax competition leads to a *race to the bottom*, i.e., it depresses the size of the public sector (see, e.g., Zodrow and Mieszkowski (1986), for an early contribution). The argument goes as follows. When a country increases its tax rate, taking the other countries' taxes as given, capital flies from the country. This increases the elasticity of the tax base vis-a-vis the closed autarkic situation, and the country optimally responds by setting a lower tax rate. The fiscal externality operates through the net return to capital mechanism. Actually, when one country increases its tax rate, the capital becomes more costly in that country, hence its demand decreases, and so does the world demand. The price of capital in the world capital market then decreases for the market clearing condition to be respected. This makes capital cheaper to the other countries. In other words, when a country lowers its tax rate, it increases the tax base of the other countries, but it fails to take this effect –*the fiscal externality* – into account. This leads it to set a lower tax rate than it otherwise would.

The welfare consequences of such a race to the bottom depend on whether one sees the government as a benevolent welfare-maximising entity or as a Leviathan (Brennan and Buchanan (1980), Edwards and Keen (1996) seeking to maximise tax revenue at the expense of the citizens' welfare. In the former case, fiscal competition is efficiency-worsening, while in the latter it is efficiency-enhancing.

The view that tax competition is efficiency-enhancing in the presence of non-benevolent governments is partially challenged by Besley and Smart (2007). These authors take a political economy viewpoint and model a world where both benevolent and non-benevolent politicians co-exist. Welfare depends on two important mechanisms: on the one hand, the voters want to *select* benevolent politicians, i.e., oust the non-benevolent ones, should they be elected in the first place and, on the other hand, *discipline* non-benevolent politicians during their periods in office. It turns out that the two objectives conflict with each other – the more disciplined is the non-benevolent politician, the less likely it is that the voters identify him and vote him out of office. Since tax competition increases the marginal cost of public funds (it is an example of the *fiscal restraint* modelled by the authors), it is helpful for disciplining purposes, but not for selection ones.

When it comes to benevolent governments, it is important to define the benchmark against which one observes the race to the bottom. Usually, one has in mind full centralisation of tax decisions, resulting in a uniform tax rate across all countries (in a world of symmetric countries). Some authors consider instead the benchmark of no capital trade. In fact, early contributions to the literature analyse tax competition between symmetric countries. In these cases, closing the borders to capital trade yields the same outcome as centralised decision making. However, with symmetric countries there are no gains to be realised from trading in capital and thus opening the borders only entails the negative consequences of tax competition.

4.1.2 Agglomeration economies - a race to the top?

The race to the bottom result relates the size of the public sector to economic integration, defined as the liberalisation of capital markets. One may equivalently define economic integration as the liberalisation of consumption goods trade, or a decrease in the level of transportation costs. Taking this viewpoint, a more recent strand of the literature (Ludema and Wooton (2000), Baldwin and Krugman (2004), Kind et al. (2000)) challenges the view that economic integration necessarily leads countries to depress the size of the public sector. In a friction-less world, the mobility of factors leads to the arbitraging of net returns across locations, thus production factors respond to any marginal change in the tax rate. On the contrary, the combination of increasing returns and costly trade may generate endogenous agglomeration of production factors in a single location (see Chapter 1 of this document). In such a core-periphery equilibrium, equalisation of factor returns across locations does not obtain. The superior return obtained in the core is sometimes referred to as the "agglomeration rent". Therefore, the firms in the core region do not respond to a marginal change in the tax rate by relocating to the periphery. This feature allows the core region to tax away the agglomeration rent. Importantly, for a whole range of trade costs, decreasing them increases the agglomeration rent. In this case, deeper economic integration actually leads the core regions to set a higher tax rate.

The possibility of a race to the top says nothing about its desirability –it is not because taxes are higher that it is welfare improving to have tax competition in the presence of agglomeration externalities. Baldwin and Krugman (2004) do show that tax harmonisation actually makes both regions worse off, if the common tax rate falls between the two equilibrium ones. Actually, such a tax rate is not low enough to attract firms for the periphery, hence its sole effect is to decrease fiscal revenue in both countries. As an alternative welfare-improving policy, the authors purpose a tax floor, where the floor is set just below the periphery's tax rate. This would not change the periphery's tax rate, but it would allow the

core to increase its tax rate. The reason is that the core sets the tax rate which makes the periphery indifferent between setting a lower tax and attracting all the firms, or setting its equilibrium tax rate and remain the periphery. By ruling out the possibility for the periphery to attract the firms, the tax floor eliminates a constraint for the core. Hence, the reform can make the core better off, without harming the periphery. The identification of a weakly Pareto improving reform clearly indicates that the tax competition equilibrium is not efficient. We shall return to the efficiency of tax competition with agglomeration externalities on Section 4.3 below.

4.1.3 Asymmetric taxation and productive inefficiency

There is an additional concern with tax competition, besides the potential *race to the bottom*, which is that of *productive inefficiency*. The advantage of capital mobility is that the equilibrium capital allocation equalises marginal productivities across locations and maximises production. This equalisation is reached through the arbitraging of different productivity levels by capital owners. In the presence of capital taxation, capital owners arbitrage the net return, which is the marginal productivity (gross return to capital) net of the tax rate. The two are equivalent only to the extent that capital taxation is homogeneous. However, the assumption of symmetric countries is not empirically reasonable. Several authors have studied asymmetric countries (see, for instance, Wilson (1991), Bucovetsky (1991), and Peralta and van Ypersele (2005), for differences in capital and labour endowment), obtaining different equilibrium tax rates. In Wilson (1991) and Bucovetsky (1991), regions have identical per capita endowments of capital, but they differ in population size. The large region has more market power in the world capital market (because its demand is a larger share of the world demand), hence the net return to capital decreases more with the large than with the small region's tax increase. The total cost of capital in the region, equal to the sum of the decreased net return to capital and the increased tax rate thus increases less than in the large region, which mitigates the capital flight effect. This mechanism ensures that the large region's cost to increase the tax rate is lower than the small region's, which leads the former to set a higher capital tax rate. The small region thus finds itself with a higher level of invested capital – a result which is known as the *small region advantage*. James and Gordon (1994) and Peralta and van Ypersele (2005) let the per-capita capital endowments of countries differ, and show that the country with the highest endowment, who ends up exporting capital in equilibrium, sets a capital subsidy, while its competitor sets a capital tax. The reason for this is that the capital exporting country gets a transfer from the rest of the world which is increasing in the net return to capital. It therefore has an incentive to lower the tax rate, as a means to inflate the net return to capital; conversely, the capital importing country rents capital from foreign citizens and wants to depress the price at which it does so. This is achieved by a high tax rate.

Low-tax jurisdictions have more production factor than high-tax ones, hence marginal productivities are not equalised. If one unit of the production factor is moved from the low-tax jurisdiction to the high-tax one, overall production increases. While the underprovision problem is of a redistributive nature, the productive distortion affects the actual amount of resources available to the economy. It hampers one of the classical benefits of market integration, i.e., improved factor allocation.

Asymmetric taxation may also result from a setup with ex-ante identical countries. For instance, Wilson (1987) incorporates capital taxation into a classical trade model with two goods and two production factors, where region are ex-ante identical. In equilibrium, the regions producing the capital-intensive good

set a lower capital tax rate, while the regions producing the labour-intensive good choose a higher capital tax.

4.1.4 The tax mix

The distortions mentioned so far are related to the tax *level*. Some papers (e.g., Bucovetsky and Wilson (1991) have looked at the tax mix, i.e., the extent to which competition for a mobile factor leads the countries to choose a combination of tax rates different from what they otherwise would. In a model with endogenous savings and labour supply, where governments set both a labour income tax and a capital tax, the authors show that tax competition shifts taxation away from the immobile (labour) into the mobile (capital) factor.

4.1.5 Empirical evidence

Empirical tests of tax competition suffer from two main problems (Wilson and Wildasin (2004)). Firstly, finding evidence in favour of tax competition says nothing about its efficiency properties, i.e., if it is good (i.e., with Leviathan governments) or bad (i.e., with benevolent governments). Secondly, regions may react to each other's tax rates for reasons other than tax competition. For instance, even if production factors are immobile, politicians may react to the competing regions' fiscal policies if voters use comparative performance to decide about re-election (the so-called yardstick competition argument).

Devereux et al. (2008) analyse tax competition among 21 OECD countries between 1982 and 1999. The empirical approach in Devereux et al. (2008) has several advantages. On the one hand, it estimates directly tax reaction functions (i.e., how a country's tax rate depends on a number of factors, including the other countries' tax rates).[1] On the other hand, its tax measures are based on careful analysis of each country's legislation, and how it would treat hypothetical investment scenarios. The authors address the yardstick competition issue by controlling for the existence of capital controls in the country (which decreases the mobility of the tax base, hence the case for tax competition). They find that strategic interaction is much stronger in the absence of capital controls.

Devereux et al. (2008) put forward a theoretical model with multinational firms (whose location is given) who decide on the investment to undertake in each country, depending on the effective marginal tax rate, and how much profit to shift from the high to the low tax country, depending on the statutory tax rate. The authors findings strongly suggest that countries compete over the statutory tax rate rate, that is, they seem to compete for firms' profits. Conversely, Devereux et al. (2008) do not find strong evidence of competition over effective marginal tax rates, a finding which goes against the empirical relevance of the classic tax competition models outlined above.

Moreover, they show that their empirical findings may explain the observed decrease in statutory capital tax rates over the period they analyse.

[1] Other studies focusing on country-level data, such as Krogstrup (2003) use aggregate measures such as total expenditure or the ratio of total capital tax revenue to GDP. The direct study of tax reaction functions has been undertaken in the context of local business tax competition by a number of authors.

4.2 The expenditure side: public infrastructure

The literature on the "race to the bottom" is mainly concerned with the underprovision of public goods that benefit consumers. However, governments provide many services which are targeted at firms rather than consumers: for instance, transport or IT communication infrastructures, or a business-friendly legal framework. The introduction of public infrastructures raises two natural questions. Firstly, are these provided at the efficient level when the governments use them as a means to attract firms, even when they are funded with a lump-sum tax? And secondly, what are the implications of allowing capital taxes to (partially or totally) fund a public good that benefits the firms? The main argument is simple and has its roots in the Tiebout hypothesis: firms may be ready to pay higher taxes if that grants them access to a superior basket of productivity-enhancing public goods.

4.2.1 Wasteful competition in public infrastructure?

The first paper to address this issue was Keen and Marchand (1997), who show that competition for mobile capital leads the governments to provide an inefficient mix of public goods, that is, over-provide infrastructure and under-provide public goods which benefit consumers. The main concern of the paper is still the underprovision of consumption public goods, except that this comes about as the result of a distortion in the mix of public expenditures, rather than a race to the bottom in the capital tax. Importantly, in Keen and Marchand's set-up there is no actual competition amongst regions, since they are considered small regions who take the net return to the mobile factor (capital) as unaffected by their policy choices.

Keen and Marchand's model features two types of public goods. One benefits the immobile residents, while the other enters the firms' production functions. The public input increases the marginal productivity of immobile labour and mobile capital, and its availability generates positive profits, which may or not be taxed by the government. In addition, the government may use a labour and a capital tax. It this setting, increasing the public input has several effects. The first is an inflow of capital, which increases capital tax revenues, whenever capital is taxed. This only happens when the firms' profits are not fully taxed however; when such a possibility exists, a standard optimal taxation argument allows the authors to establish that capital goes untaxed at equilibrium. The two remaining effects are present even in the absence of capital taxation: profits and the wage increase.

Naturally, the opposite effects take place in the other countries – this is the negative externality of public input provision, which explains its over-provision in equilibrium. The authors establish that welfare would increase if, holding tax rates fixed, expenditure in infrastructure decreases and expenditure in the public goods benefiting immobile residents increases.

It turns out that the result of over-provision of public infrastructure shows up in quite different setups. Moving away from the classical concave production function world studied by Keen and Marchand (1997), Bucovetsky (2005) analyses the provision of public inputs in a model with external economies of scale. Regions compete with each other to attract a mobile factor by providing public inputs, funded with a lump-sum tax. The mobile factor in this setting is better understood as skilled labour (as opposed to unskilled immobile labour), since the governments care for the income of the mobile factor employed in the country, rather than the one accruing to the country's residents, an assumption which is more

natural if it concerns workers.[2] The authors model an economy with three goods: a numéraire, the public input, and the differentiated good, this latter being produced under constant returns, using skilled labour, land, and the public input. Each region produces one variety of the differentiated good, which is used as an input in production at the national (i.e., encompassing all the regions) level. Put otherwise, total output is a constant elasticity of substitution index of regional output, and the optimal allocation of resources and vector of public inputs maximises this index, net of the cost of public inputs. There is thus a benefit of variety in the production process, since national production increases the greatest the number of varieties used. This acts as a dispersion force in this model, for it favours production in several regions, simultaneously. The country in itself is assumed small with respect to the rest of the world, hence the regions are competing with each other but isolated from foreign competition.

In addition, firm-level production of the differentiated good increases in the total amount of skilled labour employed in the region (call it industry size), i.e., there are external economies of scale due to, e.g., knowledge spillovers. These spillovers create an agglomeration force which may be stronger or weaker, depending on the elasticity of firm production with respect to the industry size and to the mobile factor. The situation where agglomeration linkages are so strong that the only outcome is full agglomeration of the skilled labour in a single region (akin to a black-hole in traditional economic geography models) is ruled out by assumption. The efficient outcome in this model is the result of the trade-off between this agglomeration force and the "benefit of variety" dispersion force, and it entails providing public inputs in a single region when agglomeration forces are strong, and an equal amount of investment in all regions otherwise.

It turns out that when full agglomeration is the efficient outcome, it may happen that the Nash equilibrium of the game between competing regional governments entails a subset of the regions providing positive levels of public input. This is because regional governments aim at maximising regional production, net of the public input cost, and thus engage in wasteful competition for the mobile production factor. This happens for intermediate levels of the agglomeration forces. Very strong agglomeration forces restore optimality (i.e., in equilibrium only one region provides public inputs). For weak agglomeration forces, in turn, the Nash equilibrium yields symmetric provision of the public input, but it may be greater or smaller than the efficient one. There is, therefore, concern for wasteful investment in infrastructures, and scope for efficiency-enhancing coordination amongst regions aiming at a lower overall provision of public inputs.

4.2.2 Public input provision with agglomeration economies

While Bucovetsky (2005) is the first paper to analyse the optimality of public investment in infrastructures in a model with agglomeration externalities, its setup departs significantly from the usual new economic geography one: economies of scale are external to the firm, and there are no transport costs. One question which may arise is the extent to which the results carry through to a more standard economic geography framework – Fenge et al. (2009), in a recent paper, undertake this task.

Contrary to Bucovetsky (2005), these authors show that public infrastructure need not be over-

[2]Borck et al. (2007) put forward a model of public expenditure composition in the spirit of Keen and Marchand (1997), but where the two production factors are immobile unskilled and mobile skilled labour, as in the current paper. They obtain strong evidence that German local governments distort the basket of public goods in favour of those which benefit the mobile workers.

provided. As it turns out, trade in final goods is a nice mechanism for the competing region to appropriate some of the benefits from increased public inputs, since its imports become cheaper. This, together with the fact that consumer surpluses enter the government's payoff, suffices to obtain that, under some circumstances, public input provision fails to attain its efficient level since competing regions do not take into account of the positive externalities of their public input decisions.

Fenge et al. (2009) analyse a model with a traditional good produced under constant returns to scale using immobile labour and a manufactured good which is produced under increasing returns, using immobile labour, mobile capital, and the public input, which is assumed to increase the productivity of labour or, conversely, to decrease the variable production cost. Capital is mobile, but its returns are repatriated, hence the model displays the *home market effect*, but there is no circular causation leading to complete agglomeration in *a priori* symmetric regions. It is well known that the equilibrium of such footlose capital models entails the big region hosting a more than proportionate share (i.e., larger than the respective population share) of firms. When trade costs are sufficiently low, the home market effect becomes so strong that full agglomeration in the big region occurs in equilibrium. Regions maximise the welfare of its citizens, which, given the preferences of the representative consumer, equals total income, appropriately discounted by the the price index. Public infrastructure is financed with a lump-sum tax on immobile labour.

Fenge et al. (2009) show that free-riding on the other region's level of public infrastructure may occur when trade costs are sufficiently low. In this case, only one of the regions provides public input and there is complete agglomeration. The mechanism for this free-riding is the price index linkage identified above: imports allow a region to enjoy the benefits of the high infrastructure level of its neighbour, without paying its costs. When trade costs are high, imports become more expensive and have a lower weight in consumers' surplus. Regions want to have highly productive home firms in order to decrease the equilibrium prices, hence they both supply an equal level of infrastructure, and production spreads evenly across the two locations.[3]

It turns out that it is optimal for public inputs to be concentrated in a single region for a greater range of trade freeness. Put otherwise, when the process of economic integration kicks in, the regions in the decentralised equilibrium switch too late to the agglomeration outcome. For the range of trade costs under which the decentralised equilibrium coincides with the efficient outcome in terms of the spatial distribution of production (i.e., agglomeration vs dispersion), it is still not fully efficient since the level of public inputs supplied is different from the efficient one. Under the dispersion equilibrium, regions oversupply the public input, since they are eager to decrease home prices, while they undersupply in the agglomeration one. The first result is explained by the negative externality stemming from capital flight out of the neighbouring region, while the second rests on the positive externality stemming from the decreased price index in the neighbouring region. This latter positive externality also explains why there is a range of trade costs for which agglomeration is the efficient outcome and the competing regions reach a dispersion equilibrium. This insights survive the introduction of population size asymmetries between the two regions and congestion costs in the public infrastructure.

The crucial difference between Bucovetsky (2005) and Fenge et al. (2009), is the sign of the interregional externality. While Bucovetsky's model features the usual capital flight negative externality, i.e., the region

[3]For intermediate values of the trade costs, an equilibrium in pure strategies fails to exist. We shall not explore this issue further in this survey.

that increases its investment attracts capital and decreases its availability to the other region, in Fenge et al. there is an additional positive externality, thanks to the existence of trade in goods. As usual in inter-regional competition models, negative externalities lead to over-provision, while positive externalities lead to under-provision.

In Fenge et al. (2009), the public input decreases the variable production cost. An alternative is to suppose that it has an impact on the fixed production cost and check if the conclusions regarding over or under-provision carry through. This is done by Egger and Falkinger (2006), whose model differs from Fenge et al. (2009) in that agglomeration economies stem from proximity to intermediate good suppliers, rather than the home market effect.

Vertical fragmentation of production and international outsourcing are pervasive in international trade and the impact of infrastructure policy on this phenomenon is an important policy question. Egger and Falkinger (2006) introduce public input competition in a two-country new economic geography model with vertical linkages. More precisely, the final good is produced under constant returns to scale, using immobile labour and a constant elasticity of substitution index of intermediate goods, which are produced in both countries and traded with positive trade costs. Hence, intermediate goods produced at home are cheaper for the firms, hence used in higher quantities, thus boosting labour productivity.

The intermediate goods, in turn, are produced with a linear technology that uses mobile capital. Each intermediate good producer pays a fixed cost which is measured in units of the final good. This fixed cost is composed of two parts. The first is country specific and decreases linearly with the level of public inputs, eventually reaching a value of zero when public infrastructure is at its (exogenous and finite) maximum level. The second is firm specific and does not change with public input provision. Intermediate goods are traded between the two countries (which form a free trade area) but not with the rest of the world. Conversely, capital and the consumption good are traded with the rest of the world, and since the two countries are small economies, their prices are given. Given these assumptions, the inputs used in the production of intermediate goods have exogenous prices.

Increasing public infrastructure has a positive impact on the number of home producers of intermediate goods (due to lower fixed costs), and a negative impact in the number of foreign producers (because of increased competition from the home firms). Given the increased number of varieties, the demand per variety decreases. Therefore, the total outsourcing of the home final good producers decreases (less foreign firms, each with a lower demand). Conversely, that of foreign final good producers increases. Finally, given the effect of intermediate inputs on labour productivity, wages increase in the home country and decrease in the foreign one.

The governments maximise national income (i.e., labour and capital income) net of (lump-sum) taxes. The taxes are used to fund the public infrastructure, whose unit cost is country-specific, that is, countries may differ in the efficiency of public infrastructure provision. Since capital income is exogenous, the optimal provision of infrastructure stems from the trade-off between wages and taxes. Given the functional forms assumed by Egger and Falkinger (2006), it turns out that this yields either no infrastructure at all, or the maximum possible level (i.e., the one that drives the country-specific fixed cost to zero). A country's best reply to the other's public infrastructure provision depends on the provision efficiency. Very efficient countries (i.e., with very low unit provision cost) should always provide infrastructure, and, conversely, very inefficient countries should not. When the unit provision cost is intermediate, then the country should provide infrastructure if the competitor does not, and not provide if the competitor does.

Depending on the relative provision efficiency of the two countries, this opens the door to asymmetric equilibria in which one of the countries (*the core*) provides infrastructure, thus hosting a large number of intermediate good producers, which leads to low international outsourcing in the final good sector and high wages. Conversely, the periphery does not provide any infrastructure and hosts few intermediate good producers, hence its final good sector features high international outsourcing and low wages. This need not be related to the infrastructure provision efficiency – it may happen that ex-ante identical countries end up in this situation, or even that the most productive one becomes the periphery (provided that both countries have intermediate provision costs, i.e., the difference between the two cannot be too high).

The negative wage externality of public infrastructure provision (it decreases the wages prevailing in the competing country) naturally raises concerns about the optimality of the decentralised equilibrium just described. It turns out that the inefficiency can be both in the symmetry of the infrastructure provision and in its level. That is, optimum infrastructure need not be equal to zero or to its maximum level. Egger and Falkinger (2006) show, in the context of symmetric countries, that there is scope for welfare improving policy coordination in the symmetric equilibrium in which both countries provide maximal infrastructure levels, and in the asymmetric equilibrium in which one country provides the maximum infrastructure level, and the other none at all. Conversely, the symmetric no-provision equilibrium is fully efficient.

The maximal provision symmetric equilibrium is only optimal when the provision cost is sufficiently low; however, there is a range of provision costs for which the equilibrium entails maximum infrastructure provision, but the countries would be better-off by agreeing to provide a lower level of infrastructure. Basically, the countries are trapped in a prisoners' dilemma which leads them to pay too high a price to attract intermediate good producers. Egger and Falkinger (2006) show that the optimal provision in this case is of the core-periphery flavour, but with both countries providing positive levels of infrastructure, although none attaining the maximum value. The core-periphery equilibrium, which arises for intermediate provision cost, may be improved upon if the two regions agree on a symmetric, albeit positive, public infrastructure.

Egger and Falkinger (2006)'s analysis sheds light on the debate about EU regional policy. Depending on parameter values, it may happen that the optimal policy is to promote a core-periphery pattern (and redistribute income otherwise, if equity concerns are in the agenda), or to fight against the current core-periphery pattern and promote a more even distribution of infrastructure and economic activity. Moreover, there seems to be more room for concern about over-provision of infrastructure than the reverse. Naturally, one should be careful about devising policy recommendations based on the symmetric country framework.

Given the two-part structure of the firms' fixed costs, a natural alternative to public infrastructure is to subsidise firms directly (thus decreasing the firm-specific fixed cost).[4] Egger and Falkinger (2006) show that the countries are actually willing to do so. Again, there is scope for policy coordination aiming at reducing over-provision of wasteful subsidies (although a low subsidy level need not be wasteful, in the sense that the welfare increase in the home country overweighs the welfare loss in the foreign one).

[4]Actually, given the modelling assumptions, public infrastructure works as a non-rival subsidy.

4.2.3 Vertical and horizontal differentiation of public inputs

Justman et al. (2002) and Justman et al. (2005) analyse the possibility of public input differentiation. The main insight from these two papers is that both horizontal and vertical differentiation of public inputs softens subsidy (or tax) competition and allows the regions to capture a larger surplus. This idea borrows from the well-known result that product differentiation softens price competition and allows the competing firms to obtain higher profits. Both papers assume that governments decide on the type of infrastructure in the first stage, and then decide taxes (or subsidies) in the second stage. Finally, the firms decide where to locate and the profit-maximising quantity to produce.

Justman et al. (2002) analyse a model of *vertical* differentiation of inputs. Regions decide the quality of the public input they will provide and the fee (or subsidy) they charge (or pay) to the firms that use it. Firms produce using mobile labour, whose productivity increases in the quality of the public input. In addition, firms are heterogeneous in their capacity to utilise the quality of the public input. The firms which are concerned by the use of the public infrastructure represent a low enough share of the economy that it makes sense to assume that the wage rate is exogenous. Regional governments maximise the sum of the wage bill (i.e., regional employment, given that the wage rate is exogenous) and tax revenue, net of the public input quality cost.

Justman et al. (2002) show that the number of regions which provide public inputs and attract firms is bounded, which seems to be a strategic explanation for the agglomeration of firms in a limited number of locations. This bound depends on the heterogeneity of firms' productivity. The authors then proceed to analyse the equilibrium in a two-region world, assuming that the firm's productivity distribution allows at least two regions to be active in equilibrium. The vertically differentiated equilibrium entails one of the regions offering the lowest possible quality of public inputs, while the other region offers a higher quality (which can be the maximum one, provided the provision cost is not too steep). Vertical differentiation plays a very important role in this setting, for in its absence the second stage tax competition leads the regions to dissipate all the location gains by offering a firm subsidy which is equal to the wage bill. They thus find themselves with a negative surplus, once the infrastructure cost is netted out. Conversely, fiscal differentiation softens the second stage competition and allows the regions to have positive payoffs. When regions are not differentiated, their only tool to attract firms is the subsidy, leading them to engage in a prejudicial Bertrand-like competition.

Justman et al. (2005) take the viewpoint of *horizontal* differentiation of inputs. They study two countries which compete for mobile firms, which produce using mobile labour. The firms pay a fixed set-up cost which is lower, the better the matching between the firm's type and the public input type provided in the region. For instance, a software firm does better in a region with a vast network of high-speed internet. As in Justman et al. (2002), regional governments aim at maximising the sum of the wage bill and fiscal revenue, net of the public input provision cost. The authors analyse two distinct scenarios regarding the information of regional governments about firms' types.

Under complete information, Justman et al. (2002) posit that the governments decide on a menu of bids which are fully contingent on the infrastructure and the firm's types. The equilibrium entails horizontal differentiation, and each firm locates in the region that offers it the highest subsidy, net of the adjustment cost due to mismatching between infrastructure and the firm's type.The regions charge (firm) type-contingent subsidies, which need not be positive for all the firms. The ones whose types are very

close to the infrastructure provided in each region are more likely to pay fees. In addition, this equilibrium is fully efficient. In Section 4.3, we shall present another example where modelling tax competition as a full information auction leads to efficient outcomes.

When information is incomplete, in turn, the governments cannot design type-contingent subsidies and are constrained to offer all the firms the same subsidy. This subsidy depends on the difference in the infrastructure types provided by the two regions. This leads the regions to adopt maximum differentiation (that is, provide infrastructures on the two extremes of the type distribution). Although firms split efficiently across the regions, the public input supply fails to achieve efficiency, since it entails excessive differentiation. Interestingly, the regions are better off in this incomplete information case, for two reasons. On the one hand, competition with a unique subsidy to all the firms transfers less surplus to the firms. On the other hand, maximum differentiation of public infrastructure relaxes subsidy competition a lot.

Tax (or subsidy) coordination is not optimal in this setting. If regions only compete on the input dimension, we obtain the classical result of both regions providing the (same) median type of infrastructure, which is not the efficient one. However, this allows the regions to capture a higher level of surplus. Firm's productivity losses due to type mismatching are, however, greater than the regions' gain.

4.2.4 Capital and public input competition

We now look at the interaction of capital taxes and public input competition. It turns out that the insights the literature has to offer are mostly empirical. We are still lacking a theoretical model of capital tax and public input competition, that is, one which combines the two externalities together, with the notable exception of Hindriks et al. (2008), discussed in Section 4.5 below.[5] Benassy-quere et al. (2007) provide a stylised model of a small open economy deciding on its fiscal policy, taking the world net return to capital as given (as in Keen and Marchand (1997)). We shall begin by presenting its main results, and then proceed to the empirics.

Benassy-quere et al. (2007) document the diversity of tax and infrastructure levels in the European Union countries, arguing that this fact is not compatible with a "race to the bottom" story, rather with one according to which firms look at the combination of taxes and public inputs in their location decisions. They propose a theoretical model to study this bi-dimensional competition, and then test it using a database of 18 EU countries. It is worth mentioning that Zodrow and Mieszkowski (1986) allowed for the possibility that the public good fosters the marginal productivity of capital. However, productivity increases by less than the taxation cost of providing the public good, hence capital can never be attracted by additional investment. The firms in Benassy-quere et al. (2007)'s model use immobile labour, mobile capital, and infrastructure as inputs in production. Governments maximise the utility of the representative citizen, who consumes a consumption good and the public good. Importantly, the public good which enters the representative citizen's utility function is also used in the production. The authors are not interested in the composition of public spending which is the focus of Keen and Marchand (1997). The capital tax funds an exogenous proportion of the public good – it is implicitly assumed that governments have other tax instruments to raise revenue, but these are not explicitly modelled.

Increasing the tax rate in this model does not necessarily lead to capital flight since it also increases

[5] In this paper, however, the capital tax and public input provision are not linked via a government budget constraint.

public input provision, which makes the country more attractive for capital. The authors introduce a number of assumptions about the production function which together imply a hump-shaped relationship between the net return to capital and the level of public input. In other words, the marginal productivity of capital, net of the capital tax rate (which is proportional to the public input) increases with the public input for low levels of this latter, and decreases for high levels. They also show that the equilibrium level of the public input is on the decreasing part of this curve, implying the clear-cut policy implication that it is not possible for a government to attract capital with a tax increase, even if the revenue is totally spent on the public good. Therefore, in general, the public input will be under-provided when compared to the closed economy case.[6] The only case in which over-provision may occur is when the country imports enough capital that its capital tax revenues are much higher in the open economy than in the closed economy case.

The authors then proceed to an empirical estimation of their model. They use data on foreign direct investment in 18 EU countries by US firms, between 1944 and 2002. FDI is proxied by the firms' stock of capital expenditures, that is, those aimed at acquiring or improving physical capital, while public infrastructure is measured by the stock of public capital per square kilometre. The corporate tax rate is measured alternatively by the statutory or the average effective tax rate. The authors control for several other factors, including household specific public goods (logarithm of the ratio of social public expenditures to GDP or health public expenditures to GDP), economic geography controls (sum of the distances to all the remaining countries in the data set, market size of the destination country), the nominal exchange rate between the country's currency and the USD, a measure of the share of employment in the firm's sector in the destination country, to control for Marshallian externalities, and labour market variables (sector-level unit labour costs, labour market flexibility). All the variables are measured in logarithms (except for the labour market flexibility index) and the authors perform several robustness checks with lagged variables in order to deal with potential endogeneity concerns. Time, sector, and country fixed effects are also introduced alternately.

The empirical analysis points to an outflow of capital in response to a tax increase, even when its proceedings are spent on public inputs, implying that the incentive for a race to the bottom survives the use of the capital tax to finance productivity-enhancing public goods. Conversely, increases in public good provision ought to be funded by taxes on immobile factors. Public inputs, contrary to social or health expenditures, do have a positive impact on inward foreign direct investment. Hence, one should expect tax competition to distort the composition of public expenditures away from social and in favour of productive public goods, confirming Keen and Marchand (1997) result.

It should be pointed out that the conclusion that a tax increase used to fund public inputs induces a capital flight is not consensual in the empirical literature. Gabe and Bell (2004) study the determinants of location of the 3,763 new business establishments that appeared in the 129 largest (population-wise) municipalities of the state of Maine, between 1993 and 1995. Their controls include several types of government expenditure on education, non-education expenditure, wage, distance of the municipality centre to the nearest interstate highway, a marshallian-externalities agglomeration measure[7], education

[6]Note that the term under- and over-provision are commonly used in reference to the efficient allocation, that is, the one that maximises the overall sum of welfare across all the regions.

[7]Percentage of municipality's businesses in 1-digit SIC category divided by percentage of businesses in the United States in the same category

subsidies received by the municipality, population size, and several industry and county fixed effects. Importantly, no direct measure of municipal taxes is included. The authors argue, however, that the inclusion of the other items of the municipal government budget constraint (different types of expenditure and education subsidies) allows them to interpret the coefficient of each expenditure variable as the net effect of a change in the variable along with an offsetting change in the amount of per-capita municipal taxes. The empirical results point to a positive effect of non-education expenditures on the number of new business establishments. These results should be interpreted with caution, however, since both the expenditure measure and the omitted tax variable are too aggregated.

4.3 Efficient tax competition? State aids and competition policy

When production takes place under constant returns to scale, the market outcome is efficient in the absence of government intervention. Tax competition is then bound to create inefficiencies. It is well known that when the industries operate under increasing returns, there is scope for corrective taxation. A natural question is then to which extent tax competition may have a corrective role in a world of imperfectly competitive markets. In what follows, we shall follow the literature and refer mainly to state aid or subsidies. These are still tax competition settings, however, since subsidies are nothing other than negative taxes. It turns out that subsidy competition may be efficient enhancing *(i)* if the countries may design tax schedule which is contingent on the full set of admissible firms' strategies, under some conditions, *(ii)* to decrease the inefficiently high share of firms hosted by the core region, and *(iii)* to overcome inefficient locational lock-in. We shall also point out that tax competition need not be welfare improving under all circumstances in the presence of imperfect competition.

4.3.1 The possibility of a full menu of transfers to firms

The first authors to study subsidy competition were Besley and Seabright (1999). They put forward a menu-auction model in which countries bid for the investment of one or more firms in the two countries. In the two firm case, the firms decide sequentially. The countries' payoffs are contingent on the firm's investment decision, which may simply be a location one (in which case it becomes dichotomous, that is, the firm locates in either country, but not in both simultaneously). The authors give several examples which can rationalise such a payoff profile, including the imperfectly competitive assumption, i.e., the rival firms' profits decreasing in the country that hosts the firm and increasing in the other country.

The menu-auction approach supposes that the governments submit simultaneous bids contingent on the full investment profiles of the two firms in the two periods. Under this assumption, the outcome of the bidding process is efficient, in the sense that it induces the firms' decisions which generate the highest total surplus (i.e., for both countries). This is a common feature of menu auctions. The basic intuition is that allowing bids to be contingent on the full space of investment decisions ensures that the firms completely internalise the externalities that they impose upon the other firms and consumers. Naturally, the bids entail a transfer from the countries' tax payers to the firms.[8] Importantly, efficiency crucially depends on the countries' ability to commit ex-ante to the two-period bidding strategy. Basically, the

[8]Strictly speaking, the efficiency result depends on the ability of the governments to generate tax revenue in a non-distortionary way. The result still holds as long as the distortions are not too high.

countries must be able to commit not to compete against each other in the future. Without commitment, efficiency still obtains if the payoffs are separable. Separability in this setup means that a given country's payoff can be written as the sum of two components, the first related to the firms' decision at home and the second related to the firms' decision abroad. As acknowledged by the authors, this is less likely to happen in industries where clustering effects are important (e.g., high-tech ones). The efficiency result carries through if there is imperfect information about the payoffs in the case of a simple location game. However, if the firms' strategies are more complex, uncertainty about the payoffs breaks down the nice efficiency properties of menu auctions.

Which policy implications can be derived from these insights? One would like to devise a commitment mechanism for the governments. Is a ban on state aid or a limitation (e.g., a ceiling) such a mechanism? Not necessarily, for the efficient bidding implies transfers between the governments and the firms in every period, including future ones. Hence, banning state aid altogether may well rule out efficient aid. More generally, it is not obvious that the governments may make fully contingent payments that lead the firms to internalise all the externalities they impose. The authors give the example of a multinational firm which receives state aid in country A to increase its operations there, which leads it to decrease its operations in country B. This latter should then find a means to bid for the firm to keep the activity within its borders at the same level - perhaps through a tax rebate. In theory, the country who is willing to bid the most is the one that has more to gain, hence inducing the efficient outcome. In practice, however, given the non-discrimination rules prevailing in may tax codes, it is not obvious how these transfers should occur.

ICI ***** One may conjecture that introducing an asymmetry in the fiscal competition game by banning some of the country to bid for the firms would also break the efficiency result. Indeed the European ban on state aid combined with its exception for laggard region creates such asymmetry. As one of the regions doesn't bid, the firm is not able to internalize the externalities induced by its investment there. The firm's investment decisions only take the laggard region externalities into account. This may generate inefficient decision. Note that it would also decrease the equilibrium level of subsidy given to the firms. More importantly, the objective of the exception to the state aid ban is to help laggerd regions. Having that type of region being the only one able to bid increases the probability of firms investing there. The set of the localization considered by the multinational may be broader than the European Union. In that case, binding the hands of the most developed regions in the EU may not favor the laggard ones. It may give some advantage to the region outside Europe.

4.3.2 Subsidising firms in Core-Periphery economies

We now turn to the discussion of the desirability of subsidy competition when countries compete for footlose monopolistically competitive firms, in a world of increasing returns and costly final good trade. As discussed in Chapter 1, economic activity may either be dispersed across countries or a core-periphery pattern may emerge, whereby firms totally or partially agglomerate in one of the countries. Besides the obvious pricing inefficiency stemming from firms' market power, there is another source of inefficiency in these models – namely, the location equilibrium need not be the one that maximises the sum of consumer surpluses and firms' profits. It turns out that for a whole range of parameters, the *laissez-faire* equilibrium involves agglomeration while dispersion would yield a superior outcome (Ottaviano and Thisse (2002)). Ottaviano and van Ypersele (2005) show that capital tax competition may induce some firms to relocate

from the core to the periphery, thus improving upon the *laissez-faire* location pattern.

As usual in new economic geography models, a constant returns to scale sector (the traditional one), with costless trade, co-exists with a modern sector with a monopolistically competitive structure, with a continuum of firms producing different varieties of the industrial good. In Ottaviano and van Ypersele (2005)'s setting the traditional sector uses labour, while in the modern one each firm uses one single unit of capital, irrespective of the production scale. Consumers's preferences feature love for variety (that is, given income and prices, the utility level increases with the number of varieties of the good consumed). Capital is mobile, but its returns are repatriated, hence the model displays the *home market effect* feature, but not the circular causation one. Put differently, while it is true that the big market hosts a more than proportional share of firms, firms' agglomeration does not generate agglomeration of income, because it is capital, as opposed to labour, which is mobile. It is well known that the equilibrium of such footloose capital models entails the big region hosting a more than proportionate share of firms (i.e., larger than the respective population share). When trade costs are sufficiently low, the home market effect becomes so strong that full agglomeration in the big region occurs in equilibrium.

Capital tax competition in this framework has several effects. Firstly, when a country increases its capital tax rate, given the tax rate of its competitors, some capital flies from the country. In traditional tax competition models, this capital flight hurts the country, because the social value of capital, which is equal to its productivity, exceeds what the country pays for capital, which is equal to its net return, by the value of the tax rate.[9] In a monopolistically competitive model with costly trade, there are two further channels whereby capital flight impacts the countries' welfare. On the one hand, decreasing the share of domestic firms implies that consumer have to rely more on exports, thus enjoying a lower surplus. On the other hand, the remaining domestic firms face a less fierce competitive environment, which increases their profits. Secondly, it depresses the net return received by capital owners. This effect is known in the literature as the *terms of trade effect*. This impacts negatively on the welfare of the small capital-exporting country, and positively on the big capital-importing one. This last effect explains why, in equilibrium, the big region sets a higher tax, which actually amounts to a smaller subsidy, than the small one. This happens when trade costs are not so low as to induce complete agglomeration of firms in the big region. The subsidy difference decreases the share of firms in the big region, which becomes closer to the optimum one. There is also a range of trade costs for which full agglomeration is both the free market outcome and the optimum one, and the periphery is able to attract some firms with a capital subsidy in the tax competition equilibrium. In this case, tax competition is not welfare improving. The authors also show that cooperative tax setting leads to partial convergence of tax rates (that is, the tax differential across the two countries becomes smaller, but does not vanish).

4.3.3 Overcoming inefficient locational lock-in

Borck et al. (2009) look at efficiency from a different angle. Instead of identifying the efficient spatial pattern of firm location, and comparing it to the first best, they concentrate on the full agglomeration outcome but introduce ingredients in the model that make it more efficient for this agglomeration to occur in one of the regions, as opposed to the other. Indeed, one of the consequences of allowing for increasing returns in a world of mobile production factors is the multiplicity of location equilibria. This property

[9] If capital is subsidized instead, then this effect improves the country's welfare.

leads to path-dependence on the spatial pattern of economic activity (see Chapter 1). With symmetric countries, there is no efficiency-grounded reason to prefer economic activity to locate in any particular country. This does not imply, however, that the spatial pattern of industrial production is efficient, since it may happen, for some values of transport costs, that dispersion is more efficient while agglomeration obtains in the *laissez-faire* equilibrium, as shown by Ottaviano and van Ypersele (2005). This is no longer the case when countries are asymmetric. It may then happen that firms agglomerate in a country, when efficiency would require them to agglomerate in the other. The industry is then locked-in an inefficient location equilibrium.

The authors put forward a monopolistic competition model with two regions differing in market size. There are both internal (i.e., firms produce under increasing returns) and external (i.e., knowledge spillovers) scale economies. Consumers have constant elasticity of substitution preferences featuring love for variety (that is, given income and prices, the utility level increases with the number of varieties of the good consumed). As in Ottaviano and van Ypersele (2005), capital is mobile, but its returns are repatriated, hence the the big market hosts a more than proportional share of firms. In such a footloose capital framework, in the absence of external returns to scale the equilibrium entails the large region hosting the core, as in Ottaviano and van Ypersele (2005) (whether with full or partial agglomeration). However, in the presence of knowledge spillovers, and if they are sufficiently strong, it may well happen that it is the small region that becomes the core (although the core in the large region is also an equilibrium in that case). This is inefficient to the extent that there are more consumers in the large region, who must pay higher periphery prices due to positive trade costs, than in the small region, who pay low core prices. The benefits from agglomeration stemming from knowledge spillovers, in turn, are realised irrespective of the country where agglomeration of the firms takes place.

The question is then whether capital subsidies can lead the economy out of the inefficient lock-in. Borck et al. (2009) assume that the core (i.e., large) region moves first and is followed by the periphery (i.e., small region), although their results carry through to other timing assumptions (i.e., the results about the persistence of the inefficient lock-in, not on the actual level of subsidies). The question is then whether the small region is ready to defend the core by matching the capital subsidy offered by the big region. It turns out that this depends on the relative weights of workers and capital owners on the government objective function. Indeed, capital owners' gain from relocating to the big region and getting a subsidy for it comes at the expense of lower home wages. For a whole range of parameters, and as long as the workers' weight is not too high, allowing for subsidy competition breaks the lock-in and induces the more efficient location of the firms in the big region. As in Besley and Seabright (1999)'s menu-auction approach, the result hinges on the fact that the big region has more to gain from hosting the firms, hence is willing to pay more for that. Subsidies are just an indirect way for the capital owners to internalise the location externalities they impose on the remaining agents.

Taken together, these insights suggest that the traditional negative view on tax competition may have underestimated the importance of market imperfections. Indeed, market imperfections in international trade are pervasive. A recent contribution by Boutin et al. (2009), for instance, documents that multinational firms channel resources from cash-rich subsidiaries to units that face more fierce competition, which use these transfers to devise aggressive price strategies, thus limiting entry in the market. The extent to which this sort of firm behaviour aimed at limiting competition can be tamed by tax competition is still an open question, one that is bound to attract scholarly attention in the future.

4.3.4 Subsidising immobile firms to avoid exit

Finally, we point out that imperfect competition, *per se*, is not a sufficient condition for capital tax competition to be welfare improving. Martin and Valbonesi (2008) study a setting with two countries where firms compete in a Cournot oligopoly. One of the countries hosts more efficient firms than the other, in the sense that they bear a lower fixed production cost. When barriers to trade are eliminated and the two markets are integrated, the least efficient firms (i.e., the ones with the highest fixed production costs) exit the market. The equilibrium price goes down, not only due economies of scale stemming from expanded demand, as due to the decrease in the average fixed cost. One should underline the important fact that this is not a model of mobile firms, contrary to the previous ones. The countries' subsidies are not meant to attract mobile firms, but rather to rescue immobile domestic firms from exiting the market which becomes more competitive.

The country that hosts the least efficient firms may have an incentive to subsidise them in order to induce some firms to stay in the market. This may be because the government cares for foregone profits, or because some of its consumers are not benefiting from the integrated market along the integration path. Such a subsidy is obviously inefficient, since it makes inefficient firms that would otherwise exit to stay in the market. Interestingly, the authors show that the most efficient country is better-off not subsidising its own firms, since the gain to its consumers would outweigh the subsidy. This result breaks down in the monopoly case, i.e., when there is only one producer in each country. In that case, both countries have an interest in subsidising its only firm in order to make it stay in the market. This model is a simple illustration of what may explain the prevalence of state aid by some countries in the process of European integration and it carries a clear message in favour of a strict ban of these type of subsidies to declining industries. The important difference between this and the previous papers seems to be the mobility of the firms. When the subsidies are aimed at attracting mobile firms, they are bound to allow the firms to internalise the externalities that their mobility imposes. In the declining industry scenario, firm exit does not entail externalities and the subsidies simply distort an efficient market mechanism.

4.4 Tax discrimination, Multinational firms and Profit Shifting

Firms – and capital – are not born equally, and some are bound to be more responsive to international fiscal policies than others. While the literature revised in the previous sections does not allow for such differences, there are numerous real life examples of preferential tax regimes targeted at multinational or foreign firms, presumably prompted by the (higher) elasticity of their investment or location decisions to corporate tax rates. The European Union has adopted a *Code of Conduct* of business taxation European-Communities (1998) which deems such practices as harmful and tries to prevent member countries from implementing them. We now survey the relevant literature to shed light on this issue. As it turns out, there is no consensus on the harmfulness of such practices.

One may distinguish two types of mobility. On the one hand, the actual production activity may be more mobile in some cases (that is, some types of capital are more footlose than others, perhaps because of their owners' awareness about international markets). On the other hand, multinational firms have various forms of moving the tax base from high tax into low tax countries, without changing the location of production activities. These include the option to finance an affiliate with debt or equity, the

organizational form (e.g., to own the affiliate or to engage in a joint-venture with a local firm), the payment of management fees or royalties between the parent company and its affiliates, or the manipulation of transfer prices (i.e., the prices that are used for intrafirm international trade in goods and services).[10]

Policy-wise, there are two ways in which countries may implement an effective lower rate of taxation for multinational firms. The simplest one is to set different tax rates for different types of firms. The second is to set a unique tax rate, but allow the multinationals to engage in tax planning aimed at decreasing their overall tax bill. We shall begin by reviewing this latter.

4.4.1 Only one tax instrument: indirect ways to discriminate

Without explicitly analysing the desirability of, or the decision about, tax discrimination, Hong and Smart (2010) and Slemrod and Wilson (2009) reach opposite conclusions as regards the welfare impact of the ability of multinational firms to engage in tax planning (that is, shift profits out of high and into low tax countries).

Hong and Smart (2010) model a small open economy (that is, the world net return to capital is taken as given) with two sectors, one using labour and entrepreneurial capital, both immobile, and the other using immobile labour and mobile capital. This latter is called the multinational sector. The government chooses a corporate tax rate, common to both sectors, to maximise a weighted utilitarian social function which attaches a higher weigh to the workers than to the entrepreneurs. In other words, the capital tax is an indirect way to redistribute from the entrepreneurial to the working class. The authors show that when multinationals have the opportunity to decrease their tax bill by lending an exogenous amount to an affiliate located in a tax haven, the optimal corporate tax rate of the small country increases.[11] The basic intuition is that is allows the countries to set a higher corporate tax rate, which falls on other firms besides the multinational ones, and thus increase fiscal revenue. Were it not for the possibility of decreasing their tax bills through tax planning, multinational firms, or the capital they employ, would flee the country in response to a high corporate tax rate, thus making it optimal for the country to choose a lower tax rate.

Firms decision on whether to become mobile

Slemrod and Wilson (2009) model tax competition among symmetric countries. Their setting differs from Hong and Smart (2010) in a number of other ways. Firstly, there is only one production sector, which employs mobile capital and immobile labour, but the firms may decide to pay an idiosyncratic cost which allows it to shield some of its tax bill (one may think of this as the firm starting up an affiliate in a tax haven). Secondly, tax shielding activities are sold for a given unit price by perfectly competitive tax havens, who are otherwise silent in the model (they do not produce or trade, and their tax rate is zero). Weighing the gains from tax concealment against its cost, the firms with the lowest idiosyncratic cost to become multinational actually do so.Thirdly, the tax havens are small compared to the remaining countries, thus capital flight to these destinations does not have an impact on its return in the competing countries. Fourthly, the government uses the corporate tax revenue to fund a public good and its objective is to maximise a well behaved utility function depending on the consumption and the public good. In

[10] See Hines (1997) and Hines (1999) for comprehensive surveys of the empirical literature, or Clausing (2003) for a more recent contribution.

[11] As long as the tax is smaller than 50%, an empirically reasonable assumption.

this world, increasing the capital tax increases the use of tax havens by firms and the proportion of firms who participate in tax havens. The authors show that the elimination of tax heavens is welfare improving for two reasons. On the one hand, it eliminates the waste of resources used for income shifting. On the other hand, it allows the countries to increase its public good provision, which is under-provided due to tax competition.

The crucial difference driving the results is that in Hong and Smart (2010) the degree of mobility is exogenously given, while in Slemrod and Wilson (2009) all the firms may become tax haven users (at a cost). In the absence of tax havens, firms have no way of shielding their tax bill.[12] Hong and Smart (2010) do not analyse the potential elimination of tax havens; rather, the question is whether the firms are allowed to do some tax planning. If this does not happen, capital flies from the country, since it can always earn the net return prevailing in the rest of the world. Together, these results seem to suggest that, from the viewpoint of a single country, it may be profitable to offer tax avoidance opportunities to multinational firms. However, from a global welfare viewpoint, tax havens are wasteful.

In Hong and Smart (2010), the possibility of tax shielding is exogenous for both the government and the firms, while in Slemrod and Wilson (2009) it becomes endogenous for the firms, but the governments still do not have any influence on it.Peralta et al. (2006) endogeneise the government's decision to allow multinational firms to engage in tax shifting.[13]

Governments' decision on whether to monitor multinational firms

Peralta et al. (2006) set up a three-stage game, where governments decide whether to monitor the multinational's profit shifting activities in the first stage. In the second stage, they set the tax rate, which falls both on a totally inelastic domestic tax base and on the multinational firm. In the third stage of the game, the firm decides where to locate its production facilities. There is a slight asymmetry between the two countries: the firm locates in the country where its net profits are the highest and, upon indifference, prefers one of them (call it the preferred country). The basic idea is that being tougher on the enforcement of transfer pricing rules is costly if multinational firms respond by delocalizing. A country may be appealing to multinational firms simply because it offers them a greater latitude for tax planning: even with a high profit tax, it may attract the firm by committing to be loose on profit shifting monitoring.

The firm serves both markets; however, the price is lower in the country that hosts the production plant, due to savings in transport costs. The multinational firm attempts to minimise its tax bill by shifting profits from the high into the low tax country. Its ability to do so depends on the monitoring put in place by the high tax country. The country has two strategies: it either allows the multinational to shift all its profits (loose) or none at all (strict) to the low tax country. Both countries decide on a monitoring level; in the second stage, when they set tax rates, it turns out that only the high tax country strategy will matter for the firm. The governments maximise the sum of consumer surplus with the tax proceedings. In this setting, both countries want to undercut each other's tax rate to attract the multinational. However, they do not do so indefinitely, since they are better off with a high tax rate falling only on the domestic

[12]The assumption of symmetric competing countries ensures that the firms cannot find a lower tax by moving to another country.

[13]The two first papers to endogeneise tax regulations for multinational firms take the location of the firm as given and suppose that the firm cannot cheat upon the government's regulation. In this context, governments set transfer pricing rules which lead to excess effective taxation and depressed international trade – a race to the top (Raimondos-Moller (2002), Mansori and Weichenrieder (2001))

tax base than with a very low (potentially very close to zero) tax rate that falls on both the multinational and the domestic firms. How far each country is willing to go in undercutting its competitor determines its aggressiveness in the second-stage tax competition subgame, and depends on the monitoring policies implemented in the first stage. Indeed, these determine whether countries are competing for the firm's profits or location.[14]

The main result of the paper states that at most one country – the preferred one – is loose. Hence, there is no run to the bottom, nor a run to the top, on this policy instrument. The authors show that the monitoring policy may serve one of two purposes: attracting the firm, or controlling the aggressiveness of the competitor in the tax-setting second stage. It turns out that it pays to be loose only when trade costs are high. When the preferred country is loose, it makes the other country compete for the firm's profits, while being strict makes it compete for the firm's location. The firm's location matters more for consumer surplus when transport costs are high, since this increases the import prices that the consumers would pay should the firm locate in the other country. On the other hand, the firm's profits are higher at low levels of transport costs. Hence, when transport costs are low, the preferred country chooses a strict monitoring policy, to make the non-preferred country compete for the firm's location rather than its profits in the second stage, which makes it less aggressive. Conversely, when transport costs are high, it chooses to be loose, thus attracting the firm and, moreover, making the non-preferred country compete for its relatively lower profits in the second stage. One of the implications of these results is that economic integration will eventually eliminate firm's tax planning activities, since the countries are bound to become stricter.

Firms' mobility decision and governments' monitoring decision put together

Bucovetsky and Haufler (2008) put together the country's decision to discriminate with the firm's decision to become mobile. More specifically, they study a four-stage game, which begins by the countries' decision on the tax discount offered to mobile firms. More specifically, the governments allow the mobile firms to engage in tax sheltering in such a way that their effective tax rate is a fraction of the one falling on domestic firms.[15] In the second stage, as in Slemrod and Wilson (2009), the firms decide on whether to pay an idiosyncratic cost to open an affiliate abroad, i.e., become mobile. The governments then decide tax rates, and finally the mobile firms decide where to produce. Capital tax revenues are used to fund a transfer to the representative consumer. Governments maximise the consumer's disposable income (i.e., the wage, net capital returns, and the transfer). When the governments improve the preferential regime in the first stage, they are actually softening tax competition in the second, since mobile firms are partially waived from paying taxes. The effect is akin to Peralta et al. (2006)'s aggressiveness one and drives the paper's main result. This allows them to set a higher statutory tax rate in the third stage.[16] In this setup, a coordinated decrease in the tax preference (that is, increasing the ratio between the effective tax rate paid by mobile firms and the statury tax rate prevailing in the country) is not desirable, because it induces the country to compete aggressively in the tax-setting stage.

[14]The authors show that, unless one is willing to impose rather restrictive conditions on the size of the tax bases, the equilibrium involves the countries mixing over an interval of tax rates.

[15]One may argue that there are two tax rates; it should be noted, however, that at the tax setting stage the governments have already decided the ratio of the two, so that they are effectively deciding only one tax rate.

[16]It may happen that the resulting equilibrium tax rate is so high that capital gets a negative net return. The countries then set the constrained tax rate equal to the highest possible compatible with a positive net return to capital. We shall not treat this case here.

4.4.2 Two tax instruments: direct discrimination

The possibility of asymmetric discrimination: a prisoners' dilemma

We now turn to the papers which have explicitly analysed discriminatory taxation between firms with different degrees of mobility, and whose main concern is the desirability of such practices. The earliest paper that analyses this issue is Janeba and Peters (1999). The authors model a two-stage game between two countries, where in the first stage the governments decide on whether to discriminate, and in the second stage set one (if they do not discriminate) or two (if they do) tax rates, with the objective of maximising tax revenue. The authors do not explicitly model the tax planning decisions of firms. Rather, they posit that there is one tax base which does not respond to the foreign tax (call it immobile) and another one which locates in the country with the lowest tax (call it mobile). The revenue level from each of these two bases may be bell-shaped or increasing. In the non-discriminatory case, the same argument as in Peralta et al. (2006) – namely, that the countries are better off with a high tax rate falling only on the immobile tax base than with a very low one that falls on both – applies. The authors impose the conditions on the relationship between the two bases to ensure that a pure strategy equilibrium exists. In such an equilibrium, one of the countries sets the revenue-masimising tax rate on its immobile tax base, and the other sets a tax rate just below and attracts the mobile tax base. They show that the first stage of the game is a Prisoners' Dilemma: although the efficient outcome is that both countries do not discriminate, discriminating is a weakly dominant strategy, and the game has three Nash equilibria in which at least one country plays the discriminating strategy. The reason is that when both countries discriminate they bid each other down to a zero tax rate on the mobile tax base. When only one discriminates, on the other hand, it may attract the mobile tax base with a positive tax rate. Janeba and Peters (1999) is the only paper to endogeneise the discrimination policy, and is thus the only one allowing for asymmetric discrimination policies. Interestingly, Janeba and Peters (1999) show that the equilibria in which only one country discriminates are more efficient than the one where both do. The remaining papers simply compare the tax setting game when both countries discriminate, with that in which they do not, and state the relative merits of the two alternatives.

Different degrees of tax base mobility: the importance of home attachment

Keen (2001) is the first author do so. His analysis differs from Janeba and Peters (1999) in that both tax bases are mobile, i.e., depend on the tax difference between the two countries, albeit with different elasticities. Importantly, the author assumes that the taxes vary smoothly with the tax difference, even when it vanishes, i.e., contrary to Janeba and Peters (1999), the tax base does not shop around for the lowest tax base. As in Janeba and Peters (1999), governments are revenue maximisers.The author shows that allowing for discrimination increases the tax revenue of both (symmetric) countries. The reason is that it allows the countries to set a very high tax on the less elastic tax base. The crucial difference with regard to Janeba and Peters (1999) is that there is no undercutting incentive, since none of the tax bases moves discontinuously to the low tax country. Hence, there is no risk of a total erosion of the tax revenue from one of the tax bases, as in Janeba and Peters (1999) discrimination setting.

A generalisation of Keen (2001) by Janeba and Smart (2003) shows that this result crucially depends on the overall exogeneity of the tax bases. Janeba and Smart (2003) suppose that the overall amount

of the tax base responds to the countries' tax rates, e.g., due to endogenous savings, which opens the possibility for the countries to optimally set a higher tax rate on the most mobile tax base. In equilibrium, the tax base with the highest tax is that which is less elastic with respect to the country's tax rate. This is not the same as the elasticity of the tax base with respect to the foreign tax rate, which determines the tax base's degree of mobility. Were it not for the endogeneity of the overall size of the tax base, the two elasticities would coincide (that is, the most responsive to the home tax rate is also the most responsive to the foreign one).

Haupt and Peters (2005) also analyse a setup of revenue-maximising countries with two tax bases of different elasticities. However, they assume that the two countries have asymmetric access to the two bases, i.e., each tax base has a home bias for one of the two countries. Specifically, they suppose that when the two tax rates are equal, one of the bases is mostly (i.e., more than half) invested in one of the countries, and the other is mostly invested in the competing country. In a way, each country has a domestic and a foreign tax base, while Keen (2001)'s distinction is between a more and a less mobile tax base for *both* countries. This alternative assumption changes the efficiency implications of restricting discrimination sharply: total fiscal revenue is maximised when the countries are forced to setting the same tax rate on both tax bases. They also show that a partial ban on tax discrimination, i.e., imposing a maximum difference between the tax rates applied to the two tax bases, increases fiscal revenue. This partial ban implies that tax discounts offered to the foreign tax base also apply to the domestic one. Countries are hence less eager to offer tax discounts – as in Peralta et al. (2006) and Bucovetsky and Haufler (2008), what matters is that tax competition is softened.

The crucial difference between Haupt and Peters (2005) and Keen (2001) lies in the fact that in the latter the more elastic tax base is more elastic for both countries. In the former, on the contrary, the more elastic tax base for one of the countries (the foreign one) is the less elastic (the domestic one) for the other country. This advantage that each country has over its domestic tax base can be exploited by appropriate coordination policies as a means to increase the overall tax burden.

The desirability of restricting preferential regimes is ultimately an empirical question about what exactly differentiates tax bases. In any case, the policy implications one can take from the literature are limited by the fact that it mostly relies on symmetric country settings. The fact that countries implement asymmetric policies, both regarding the tax rate and the monitoring of multinationals' tax shielding activities, has been documented empirically (see, e.g., Bartelsman and Beetsma (2003)), suggesting the need for further research relying on asymmetric settings.

4.5 Equalisation grants

Equalisation schemes aimed at reducing fiscal imbalances across jurisdictions exist in most federations. These schemes vary in form and complexity, but they are mostly based on one of two types: tax revenue and tax base equalisation. According to the former, the transfers to a given region are equal to the difference between its per capita tax revenue and the average per capita tax revenue of the remaining regions in the federation. The German interstate transfer system is mainly driven by this principle (Kothenburger (2002)). Under the latter, also know as the Representative Tax System (RTS) equalisation scheme, the transfer is equal to the difference between the region's per capita tax base and the average per capita tax base in the federation. One example is the Canadian intergovernmental transfer scheme

(Kothenburger (2002)). In addition, the schemes may provide full or partial equalisation. The former entails all the regions having the same tax base (or revenue). Under the latter, the scheme implements a resource transfer from the regions with the highest tax base (or revenue) to the ones with the lowest, in such a way that the gap among the regions is decreased, but not eliminated.

These schemes' primer objective is to grant the citizens across all jurisdictions an equalised access to local public goods. However, bearing in mind that each region receives a transfer which is related to its own fiscal base or revenue, and hence to its policy choices, one may expect the regions to react to the schemes by adjusting their policy choices. The fact that regions react to transfer schemes can actually be used to enhance the outcome of tax competition. Wildasin (1989) and James and Gordon (1994) show that appropriately designed transfers – the so-called *matching grants* – from the federal government lead the regions to implement the efficient taxes. In other words, federal transfers to the regions may be designed in such a way that the regions internalise the fiscal externality and the tax competition equilibrium is no longer inefficient. The transfers act as a Pigouvian subsidy which gives the right incentives at the margin for the regions to set efficient taxes. The federal transfer schemes put forward by these papers are not, however, concerned with the correction of fiscal imbalances.

A natural question is then if the two objectives (equity and efficiency) can be reconciled and, relatedly, whether one or the two types of equalisation grants observed in reality (revenue or base equalisation) may serve the purpose of restoring efficiency of the tax competition equilibrium. As shall become clear from the papers reviewed below, the literature is quite optimistic about this possibility. It turns out that under some conditions, and in a variety of settings, appropriately modified equalisation grants improve upon the fiscal competition outcome, sometimes completely restoring efficiency.

The first papers to study fiscal equalisation show that it can induce the efficient regional distribution of population (that is, with mobile labour, rather than capital) (see, e.g., Boadway and Flatters (1982)). These papers do not, however, take into account the fact that regional governments may react strategically to the equalisation scheme. In Smart (1998), the governments do react strategically, and end up setting inefficiently high tax rates. However, Smart (1998) considers immobile tax bases. When one puts together strategic tax setting with the mobility of the tax base, it turns out that Boadway and Flatters (1982)'s insight in favour of equalisation grants is restored. We now survey a number of papers that show this, in different contexts: capital tax competition with and without agglomeration externalities, profit tax competition, and public input competition.

Before proceeding, let us outline the basic intuition. In a standard capital tax competition model, there is the potential for a race to the bottom, since the regions foresee a capital flight when they increase their tax rate, therefore increasing the cost of a tax rate increase. When a given region increases its tax rate, its tax base decreases, while that of the competing regions increases (given the competing regions' tax rates). Take, for the sake of the argument, the tax base equalisation scheme. Under this scheme, the region recovers part of the lost tax base due to the tax increase. The cost to increase the tax rate is therefore lessened, and the race to the bottom mitigated. What the scheme does is to pay the region for the positive externality (higher tax base) it imposes on other regions, thus internalising the externality and moving the equilibrium closer to the efficient outcome.

4.5.1 Capital tax competition

Kothenburger (2002) sets up a model with an arbitrary number of regions with asymmetric labour and capital endowments, but with the same per-capita capital endowment. The private good is produced with a constant returns to scale technology that uses mobile capital and immobile labour. The federal government implements an equalisation scheme whereby each region receives a transfer of the tax base or tax revenue equalisation type. These scheme is budget balanced, implying that some regions actually pay, or, equivalently, get a negative transfer. Each regional government provides a public good, financed with capital taxes and the equalisation transfer, in order to maximise a well-behaved utility function that depends on private consumption and the public good. Kothenburger (2002) studies both the small region case (that is, when the regions take the net return to capital as given, hence there is no fiscal externality) and the case where regions' policy choices have an impact on the other regions' tax bases. Kothenburger (2002) is concerned with the impact of the equalisation scheme on tax levels, rather than its optimality.

Kothenburger (2002) identifies the effects of a given country's increasing tax rate on the equalisation grant received from the central government, given the policy choices of the competing regions. With tax base equalisation, the region's tax base decreases, thus increasing the transfer (the *direct effect*). On the other hand, the average tax rate and average tax base used in the equalisation formula also change, because of the change in the net return to capital (the *strategic effect*). The average tax rate is defined as the ratio of overall (i.e., across all regions) tax revenue to capital endowment. Total tax revenue increases with the region's tax rate, provided its own tax revenue does not decrease (or at least, not too much). A sufficient condition for this to arise is that regions set taxes in the upward slopping part of the Laffer curve.[17] The sign of this effect depends on where the region stands with respect to the average tax base: capital exporters are below, and get a higher transfer, while capital importers are above, and get a lower transfer. Under tax revenue equalisation, the same *direct* and *strategic* effects are present, but they differ in nature. Under the conditions above, the direct effect is negative (i.e., increasing the tax rate increases tax revenue, leading to a lower transfer) and the strategic effect is positive (i.e., the average tax revenue increases, leading to a higher transfer). Tax-revenue equalisation schemes entail a common pool effect due to the positive retention rate of region's tax revenues, which amounts to a strong disincentive to tax.

Looking at the small region case turns off the strategic effect. The results under the two schemes are then in sharp contrast: tax rates increase under tax base equalisation, and decrease under tax revenue equalisation. One should note that when regions are symmetric, there is no need to implement a system of equalisation transfers in the first place. It is still interesting, however, to note that an equalisation formula based on the tax base actually corresponds to the optimal matching grant in a symmetric region setting.

The large region case yields the following results. Firstly, symmetric regions always increase their tax rates under the tax-base scheme, and decrease under the tax-revenue one, provided that capital demand is not too elastic. Again, full tax-base equalisation schemes restore optimality. Secondly, with asymmetric regions, the most populated region sets a higher tax base than the least populated one. This result has been discussed in Section 4.1 and stems from the region's market power in the world capital market. When regions tax in the upward slope of the Laffer curve, fiscal equalisation aimed at equalising tax

[17] The Laffer curve relates tax revenue to the tax rate. Given that the tax base decreases with the tax rate, it is conceivable that the revenue decreases beyond a certain tax level.

bases always increases the most populous region's tax rate. For the least populous region, the result is less clear-cut: the tax rate increases only if the strategic effect is not too strong, or if the region has strong redistributive concerns. Under the alternative tax-revenue equalisation scheme, the strategic effect partially counteracts the negative direct effect, and it may happen that tax rates increase.

When regions are symmetric, equilibrium taxes are inefficiently low. It is thus straightforward to establish that tax-base equalisation grants are efficiency-enhancing, while tax-revenue ones are not. When regions are asymmetric, it is no longer true that tax rates are too low. There are two sources of inefficiency (public good provision and productive inefficiency, i.e., the failure to equalise the marginal productivity of capital across locations), hence we can no longer make statements about the desirability of equalisation grants based solely on the tax rate level. Bucovetsky and Smart (2006) undertake the task of studying the optimality of the transfer schemes, beyond the tax level question. Their model is similar to Kothenburger (2002), except that savings (hence, capital supply) are endogenous. Consumers receive an endowment of the consumption good in the first period, and decide how much to consume, and how much to save, which is invested as capital in production, in the second stage. In the second period, their income is equal to the sum of capital return and the firm's profits.[18] Bucovetsky and Smart (2006) derive the optimal matching grant (in the spirit of Wildasin (1989) and James and Gordon (1994)) in this context and then proceed to compare it to tax-base equalisation (whether full or partial).

Bucovetsky and Smart (2006) start by showing that with symmetric regions and exogenous savings full equalisation decentralises the optimal tax rates (as in Kothenburger (2002)) . Bucovetsky and Smart (2006) show that this result can be generalised in a number of directions. Firstly, capital demands may be asymmetric across regions, provided they only differ by a multiplicative parameter. This would obtain, for instance, in Kothenburger (2002)'s setup with symmetric per-capita capital endowments and different population sizes. Secondly, if savings are endogenous and capital demands respect the property above and are also log-concave, optimal tax rates can be decentralised with a scheme of partial equalisation, in which the fraction of tax bases which is equalised depends on the semi-elasticities of capital demand and supply (which are the same across all regions, given the assumptions on capital demand). The transfer must also include a lump-sum grant which is equal to optimal public good level. Basically, the lump-sum part ensures that the regions have enough resources to provide the public good optimally, and then the partial tax base equalisation grant gives the right incentives for efficient tax setting (hence, efficient capital allocation). Allowing for more general asymmetries across regions (e.g., public good preference, or production functions), the authors show that partial tax base equalisation achieves an outcome which is quite close to the optimal one, provided that the number of regions is arbitrarily large.[19] This boils down to the small region case, since in this case regions are atomistic in the world capital market and do not influence the net return to capital.

Egger et al. (2010) use data on municipal business tax rates in the German state of Lower Saxony and

[18] Bucovetsky and Smart (2006) assume that production uses land and capital, and no labour. The profit is then the return to the land production factor, which is equivalent to the wage if labour were used instead of land. In addition, they enjoy a public good, which is provided by the government and financed with a source-based capital tax. The federal government commits to a transfer scheme in the first stage of the game; governments then decide their taxes non-cooperatively and, finally, consumers and firms make their consumption and investment decisions.The other important difference with regard to Kothenburger (2002) is that the utility function is separable between the private and the public good, and linear in the private consumption of the second period. This implies that saving decisions do not vary with income.

[19] For the result to go through the number of different production functions must be smaller than the number of regions or, in other words, some regions must share identical technologies.

show that tax base equalisation does have a positive impact on tax rates. The particular type of tax base fiscal equalisation in this state partially compensates the municipalities for the difference in the "fiscal need", which is computed according to a formula that depends on municipal population, and the actual municipal tax base. In addition, there is a tax base floor which is guaranteed to all the municipalities, i.e., the fiscal base falling short of that floor is fully compensated for. With this scheme, the incentive to increase tax is straightforward, since it depresses the tax base and increases the transfer. The authors use a reform undertaken in 1999 that increased the equalisation rate for some municipalities and decreased it for others. Using a number of relevant controls, Egger et al. (2010) show that the former responded by increasing the tax rate, while the latter decreased taxes.[20] Buettner (2006) obtains similar findings for a panel of municipalities in the German state of Baden-Wüttemberg over the 21 years spanning from 1980 to 2000.

4.5.2 Capital tax competition with agglomeration economies

Equalisation transfers are also helpful in the presence of agglomeration economies. Gaigné and Riou (2007) introduce tax revenue equalisation in Ottaviano and van Ypersele (2005)'s two-region footlose capital model. Specifically, each regional government runs a budget balanced redistribution scheme between mobile capital and immobile labour, i.e., it taxes immobile labour and uses the proceedings to subsidise mobile capital (or the other way around). The federal government, in turn, puts in place a system of transfers aiming at partial equalisation of capital tax revenue.

In equilibrium, regions subsidise capital, with the small one setting a higher subsidy. As in Ottaviano and van Ypersele (2005), tax competition leads to inefficiently different tax rates, in the sense that there is an optimal tax wedge which is lower than the one obtaining in equilibrium, and this distortion is worsened by trade liberalisation. Gaigné and Riou (2007) show that the transfer scheme has the two effects identified by Kothenburger (2002). The direct effect leads to a tax rate decrease (i.e., subsidy increase), since the contributor wants to decrease its contribution, and the net recipient aims at increasing its transfer.[21] In addition, the direct effect is stronger for the big country, who has a larger tax base. The direct effect therefore acts in favour of tax convergence.

As regards the strategic effect, while own tax revenue increases, the other country's tax revenue actually decreases, since it amounts to a subsidy applied to a larger tax base. The effect on the average tax revenue depends on which of the two dominates. For the small region, since its subsidy is larger, the own-revenue effect dominates and the average tax base increases. For the big region, this is also the case as long as trade costs are low. In both cases, the strategic effect induces regions to increase their taxes, since a higher average fiscal revenue increases the transfer (or decreases the payment, for the net contributor). Moreover, the big region effect is weaker, which again acts in favour of tax convergence.

Gaigné and Riou (2007) show that, irrespective of the level of trade liberalisation, the tax revenue equalisation scheme decreases the tax wedge and improves upon the equlibrium outcome. Moreover, full equalisation implements the first best. The result may be extended to the case of many competing regions,

[20] The authors address the potential self-selection bias due to the splitting of municipalities into two groups with a matching procedure. They also perform a robustness test of comparing the Lower Saxony municipalities with those of other states where no reform took place.

[21] By definition, increasing a subsidy increases tax revenue, since the tax mobility actually decreases the amount paid, that is, the regions are in the upward-slopping part of the Laffer curve.

(where the strategic effect disappears), as well as to tax base equalisation.

4.5.3 Profit tax competition

Liesegang and Runkel (2009) study the impact of fiscal equalisation in a model with a multinational firm that owns subsidiaries in several symmetric countries. The firm produces in all the locations with a decreasing returns to scale technology that uses mobile capital and immobile labour and generates positive profits (equivalently, there is a third production factor like land or entrepreneurial skills that gets the rent which is left after capital and labour are paid their marginal returns). The firm may decide to shift an amount of profits out of each country, and pays a convex concealment cost to do so. In addition, the firm cannot evade taxes, i.e., the total profits shifted must sum up to zero. Each country's representative household owns an equal share of the multinational and is paid dividends accordingly. In addition, she consumes labour and capital income. Countries set the profit tax rate in order to maximise a well-behaved utility function combining a private consumption good and a public good. The public good is funded with profit taxes and equalisation transfers, which are budget-balanced and run by a federal government.

Liesegang and Runkel (2009) study the two most common ways to tax the multinational firm: separate accounting and formula appointment. Under the former, the firm is taxed according to the profits it declares in each country, while under the latter the firm's total profits are consolidated and apportioned to the countries according to a given formula. Liesegang and Runkel (2009) opt for a quite general apportionment formula entailing a weighted average of each country's share of capital, sales (or production) and wage bill in the world total.

Under separate accounting, there are several fiscal externalities induced by a given country's tax increase. As regards private income, there are three effects. Firstly, the net return to capital decreases, thereby decreasing capital income in all the countries (negative externality); secondly invested capital increases in competing countries, leading to a wage increase (positive externality); thirdly, the overall profit of the firm decreases, and so do the dividends received in all countries (negative externality). As regards fiscal revenue, the tax rate increase leads to higher outward profit shifting, which increases the tax base in competing countries (positive externality). On the other hand, one cannot determine the variation in the before profit shifting profit realised in competing countries, for it entails two positive effects (increased invested capital, hence, increased production, and decreased capital bill) with an increased wage bill. When taxation of the multinational firm respects the formula appointment principle, the fiscal externalities related to private consumption are the same as under separate accounting. The ones related to fiscal revenue are slightly modified. On the one hand, the consolidated profit increases, thanks to the depressed net capital remuneration. On the other hand, the multinational firm does not engage in profit shifting activities, since the tax bases across countries are consolidated. What it may do is distort the part of the consolidated profit accruing to a given country through the variation of the formula appointment components. It does so by depressing capital and labour demand in high-tax countries. Therefore, a tax increase generates a positive externality on the competing countries.

The authors show that a fiscal revenue equalisation scheme can only deal the two last externalities, that is, the ones related to fiscal revenue. In order to fully restore efficiency, the federal government must implement appropriately modified tax revenue equalisation, together with private consumption equalisation. In other words, each regional government should receive a transfer which is equal to the

difference between the sum of its tax revenue and private consumption and the average federation-wise value of these two variables. None of the real-world equalisation schemes actually implements something close to private consumption equalisation, hence one may expect profit tax competition in this setup to remain inefficient, despite the implementation of a fiscal equalisation scheme. One cannot either (and the authors do not take this avenue of research) claim that fiscal revenue equalisation improves upon the tax competition outcome because it solves some of the externalities, for such an argument suffers from the the usual second-best flaw that it does not suffice to count the number of externalities to evaluate the (in)efficiency of an equilibrium.

4.5.4 Public input and tax competition

When regions compete for mobile capital with public inputs, fiscal revenue equalisation allows the competing region to appropriate some of the benefits of increased infrastrucure. Since its provision is costly, this may lead the the regions to provide inefficiently low levels of public inputs. The natural question is then how does this potentially negative effect combine with the positive one stemming from alleviated tax competition, and whether fiscal equalisation is ultimately desirable in this case. This is studied by Hindriks et al. (2008). The authors model a federation consisting of two asymmetric regions. In each region, production uses an immobile factor, mobile capital and public infrastructures according to a constant returns to scale technology. Capital is taxed at the source. The cost of the public input is a convex function of its level.[22] Each regional government chooses a level of public input in the first stage, and a capital tax in the second, as to maximise the sum of the returns to the immobile factor and total fiscal revenue, net of public input provision costs. The federal government implements a partial equalisation scheme whereby each regional government transfers a given share of its revenue to the other government. The marginal productivity of capital is higher in one of the regions, granting it an advantage in attracting the mobile factor. Public infrastructures may potentially overcome this asymmetry.

Given the regional asymmetry, the marginal return to public investment in the advantaged region is higher, and efficiency requires a higher level of public input, and hence invested capital, in this region. Although the equilibrium features asymmetric provision, it fails to reach the necessary investment in the advantaged region. In addition, the less attractive region under-taxes with respect to the optimum, as a means to attract mobile capital.

If regions are symmetric, there is both under-provision of public infrastructures and under-taxation in equilibrium. The under-taxation result is the usual run to the bottom. The under-investment one may seem counter-intuitive in light of the over-provision results surveyed in Section 4.2. The regions use the first stage low infrastructure provision as a means to soften capital tax competition in the second stage, since capital is then less productive and regions are less eager to compete for it. The impact of fiscal equalisation in this symmetric setting is, not surprisingly, to reduce public input provision. However, tax rates do not change with the introduction of fiscal equalisation, since the negative direct effect and the positive indirect one identified by Kothenburger (2002) cancel out given the production and payoff functions considered in Hindriks et al. (2008). Although the only effect of fiscal equalisation is to depress further the under-provided public investment, it is nonetheless welfare improving, as shown by Hindriks

[22]The authors do not suppose a budget constraint where capital taxes are used to fund the public input. Implicitly, there are other revenue sources, and increasing budgetary requirements forces the government to use increasingly distortionary tax instruments, which may explain the convex provision cost.

et al. (2008). This is a typical paradoxical second-best result. The optimum public investment level makes sense provided that tax rates are also set at their optimal level. With under-taxation, public investments are a wasteful device to attract capital, whose return is taxed at too low a rate in the second stage.

Fiscal equalisation is also efficiency enhancing when regions are asymmetric. While there are many different effects at work in this case, this positive result is driven by the decrease in average investment, and a better capital allocation (capital relocates to the region with a productivity advantage). The fact that total welfare increases does not imply that both regions benefit from it. This is potentially an important policy question, for the rich region may resist the introduction of fiscal equalisation if it loses too much with it. Hindriks et al. (2008) show that the rich region in this setup gains from the introduction of a marginal (i.e., where the share of revenue which is transferred to the other region is not too high) equalisation scheme.

Chapter 5

Regional policy implications from both empirics and theory

From this survey, the general conclusion in terms of regional policy implications is that no simple universal recommendation stems from the academic literature.

Economic geography models lead to the conclusion that over-agglomeration may occur even under a normative criterion that consists in the maximisation of the aggregate real income. When some aversion to inequality is considered, the set of situations where this arises is extended. By over-agglomeration, we mean that there is a range of parameters such that either agglomeration takes place when it would be better to have dispersion, or that it would be better to reduce the degree of regional asymmetries while staying in a partly agglomerated equilibrium. Conversely, this means that for another range of parameters, the market delivers the "efficient" outcome, be it with agglomeration or dispersion. In the first section of this last chapter, we shall discuss this question of efficiency according to where the economy lies on the bell-shaped curve that relates trade cost and agglomeration. Concomitantly, we will discuss the efficiency gains that could be achieved by further reducing trade costs.

In a second step, our analysis will start from the general conclusion reached in empirical economic geography, which is that agglomeration delivers productive efficiency gains, reflected directly in the firms' or workers' productivity, in their location choices or in their innovative behaviour. We will discuss the regional implications of such evidence and the different ways policymakers could influence density. Lastly we will discuss the relative efficiency of the different tools policymakers can use at the local level.

5.1 The bell-shaped impact of trade costs on spatial concentration

Does the bell-shaped curve hold?

One of the main results of recent economic geography models is the presence of a bell-shaped relationship between trade costs and spatial disparities. Typically, trade integration first increases spatial concentration and disparities, due to agglomeration forces that develop more strongly than dispersion forces. At further stages of integration, the reverse holds, as dispersion forces now strengthen faster. All this has been discusses in section (1.2) and the bell-shaped relation in regional indirect utilities of a two-region economy is represented in figure (1.6).

One consequence is that trade integration is always good, for all agents, if very deep integration can be achieved. Given that the perfect integration / zero trade costs situation cannot be achieved and that in general agents are inequality-averse, there exists an optimal level of trade integration that does not necessarily correspond to the lowest level of trade costs. Unfortunately, it is characterised neither in models nor in their structural estimation. Moreover, the policy-maker wishing to implement spatial policies faces two further big problems. The first, implicit in the previous discussion, is that the optimal policy may crucially depend on where the economy is on the bell-shaped curve, since disparities first increase then decrease. The second is that it has never been fully proved that the two-region bell-shaped curve presented in figure 1 holds in the context of a large number of regions.

Regarding the first issue, if the economy is already in a situation of low trade costs where both efficiency and equity objectives are compatible, the optimal policy is clearly to further decrease trade costs. However, for higher trade costs, when the efficiency-equity trade-off holds, very inequity-averse societies would probably prefer to increase trade costs. Indeed, even when it is possible to maintain the standard of living of immobile agents when trade integration occurs (by using lump-sum transfers), some degree of disparity remains. It is therefore crucial, from the empirical point of view, to assess where an economy lies on its bell-shaped curve before being able to assess whether further trade integration is good or not.

Now we come to the second point. Given that an economy rarely consists of only two regions, one first needs to determine how the bell-shaped curve can be extended to a context of numerous regions. Extending the model to a large number of regions is fairly easy even with trade costs that are origin-, destination- and industry-specific, as they are in the real world. What proves difficult, and this should not come as a surprise given the difficulties already encountered in two-region settings, is the characterisation of the number and nature of equilibria that exist. The only possible strategy consists in appealing to simulation-based approaches, which is described in the next section.

Where are we on the bell-shaped curve?

Forslid et al. (2002) seek to evaluate the properties of an economic geography model calibrated for European regions and countries. In particular, they assess whether a bell-shaped pattern exists for regional disparities when trade costs decline. They also seek to take into account real features absent from economic geography models and they consider, among other things, the role of traditional comparative advantage effects. The economy consists of ten large regions.[1] Two sectors (agriculture and energy) use labour as the single input with diminishing returns to scale and have zero trade costs. Two other sectors (public and private services) correspond to non-tradable goods also produced under diminishing returns to scale. Finally, ten sectors operate under the standard Dixit/Stiglitz assumptions, with origin-, destination- and sector-specific trade costs. These sectors use primary factors (unskilled labour, skilled labour, capital) immobile between regions but mobile between sectors and the production functions are nested Cobb/Douglas and CES functions, with full input-output matrixes and region-specific technology. Lastly, consumer preferences also correspond to nested Cobb-Douglas and CES functions with sector- and region-specific consumption shares. Forslid et al. (2002) calibrate all the model parameters using

[1]Four Western Europe areas (Central, North, South, and West), the United States and Canada, Southeast Asia (incl. Japan), China and South Asia, former Soviet countries, Eastern Europe, the rest of the world).

national accounting data as well as other academic studies for the elasticities of substitution and trade costs. They first provide a set of simulations relating trade costs to the degree of spatial concentration in each industry.

A bell-shaped curve is observed in four sectors: those where increasing returns to scale are large. Comparative advantage effects dominate in the four others, which implies a monotonic increase in spatial concentration with trade costs. Spatial concentration variations are weaker for the former, suggesting that trade integration induces more spatial reallocation in traditional industries. Importantly, Forslid et al. (2002) conclude that most sectors with increasing returns to scale are to the right-hand side of the peak of their bell-shaped curve, i.e. in the area where reducing trade costs would further increase spatial concentration. This is also the case for manufacturing as a whole, with a peak of spatial concentration reached for trade costs 30% lower than their current value. Further trade integration would increase efficiency but also inequality and thus would not be desirable for highly inequality-averse agents.

Lastly, Forslid et al. (2002) also assess the impact of trade integration on factor returns. The conclusion is that, consistently with economic geography models, factors do not experience the same variations in their real returns with respect to trade integration, and this depends on the region. In particular, real returns may either increase or decrease, or be either bell- or U-shaped when trade integration improves. Variations are small and no large gain emerges from trade integration. Therefore, such a simulated approach using a large-scale economic geography model really provides further insights for policy-makers thinking about a possible decrease in trade costs. Gains and losses of the various agents are well identified in a framework that considers many direct and indirect effects.

A number of concerns are specific to this study, while others are more general. Taking theory seriously implies that the results obtained are valid only under the assumptions made. It is very difficult to assess the extent to which results depend on the technical assumptions made regarding functional forms for utilities or production functions, the number of sectors or inputs and even the spatial concentration index chosen to evaluate the degree of spatial disparity. As always with simulation, the only solution consists in repeating the exercise using different sets of assumptions. Importantly for the study of spatial policies, Forslid et al. (2002) assume, like Redding and Venables (2004), that no spatial labour migration takes place between countries. As discussed in Chapter 2, this clearly eliminates some channels of regional disparity, especially those that are self-reinforcing due to the endogenous size of regional population, characteristic of the Krugman (1991b) type economic geography models. Although many more general equilibrium effects are considered here than in the empirical studies presented in Chapter 3, in particular the role of endogenous demand and intermediate input prices, evaluation of the impact of a reduction in trade costs is still somewhat *ceteris paribus*, at least as regards households' location choices. This should be extended.

If a more structural approach allows researchers to be more precise about the welfare impact of trade integration, the cost of reducing trade costs still needs to be weighed against possible welfare gains. But these costs are difficult to assess and the question arises of the time horizon over which they have to be written off. The tools that can be used to reduce trade costs and the magnitude of their possible decrease given the current level of technology are related questions that are for the moment largely prospective. For many areas, including European regions, there is little to be gained from further reducing trade barriers. Information and transport costs can probably be reduced further, but by how much and at what cost are difficult to assess. In other words, whether it is possible to reach the peaks of some of

the bell-shaped curves obtained by Forslid et al. (2002), which are typically 30% below current levels, remains an open question. Assumptions about trade costs may largely limit possible inferences when, as for instance in many empirical studies, they are assumed to depend on distance only. The Redding and Venables (2004) approach, which considers components other than distance in the trade equation and which could be extended, is certainly more appropriate for distinguishing the role of trade policy and of reduced transport costs.

We mentioned earlier that dealing simultaneously with efficiency and equity concerns probably requires the introduction of a fiscal policy concurrently with trade integration. Clearly, a calibrated economic geography model could allow for that. It would tell us whether it is possible to compensate those who lose from trade integration by redistributing from those who gain. It would probably be possible to go even further and include a second-best taxation block in the model. Simulating its properties when trade costs decrease would provide further insights for policy-makers, at least under the modelling assumptions made.

Playing with simulated economic geography models certainly opens many interesting avenues for the study of regional policies. The fact that the underlying model has not been tested remains an important limit, however. Structural approaches, like those by Hanson (2005) and Redding and Venables (2004) described above, could however be used as preliminary steps in simulations. First, structural estimations could provide the values of certain parameters needed for the simulations. They could be estimated in a context fully consistent with the simulated one and not borrowed from other studies sometimes quite far-removed from the chosen framework. Second, and as stressed above, the simulated model could be tested before policy experiments are conducted, and it could at least be shown that it is not rejected by the data used. Therefore, we believe that simulating economic geography models that are structurally estimated is probably a promising line of research for studying the impact of regional policies, including trade integration. This objective has not yet been achieved. For instance, the Redding and Venables (2004) strategy may not be directly appropriate for that. The use of fixed effects in the trade equation estimation limits the simulation exercise since, by definition, fixed effects, which proxy for market size, do not change when trade costs decrease. Still, it is certainly possible to use similar settings that would allow policy-makers to address regional disparity issues, taking into account many direct and indirect effects of their decisions.

5.2 Density: what should we do about it?

Clearly, the main outcome of the empirical literature consists in the characterisation of the optimal characteristics of the regions, in terms of both overall size and industrial composition. For instance, the positive effect of density on productivity that is always obtained in such estimations clearly implies that increasing the size of regions induces productivity gains, attracts more firms, and generates more innovation.

First, however, it is difficult to evaluate the magnitude of these gains precisely. For a country like France, a doubling of density leads to a productivity gain of 2% at the lower bound, but almost 5% at the upper bound, depending on the controls introduced and, more importantly, on whether the endogenous quality (skills) and quantity (number) of the labour force are controlled for. Typically, controlling for the endogeneity of density reduces its impact by 20%, while taking individual skills properly into account

divides the estimation of agglomeration economies by a factor of 2. This makes quite a large difference for the policy-maker thinking about improving the productivity of local firms and workers by increasing density.

Such differences in estimates are partly related to an interpretational issue that is rarely commented. The 2% obtained for France corresponds to the density elasticity controlled for both skills and endogeneity. Doubling employment density increases the productivity of any worker by 2% independently of the composition of the labour force (but keeping the industrial structure constant). Now, on aggregate grounds, when one doubles the density in a region, one can simultaneously match the current regional skills composition, and the gain is 2% for all workers in the region. But one can also match the skill composition of the regions that are twice as large, biased towards higher skills. In this case, one must consider the total effect of density, i.e., not only the 2% corresponding to the direct effect, but also the extra gains due to the presence of more skilled workers, resulting in a total increase in regional average productivity of around 4%. Since density is endogenous, the policy-maker can also expect increased productivity to attract even more people to the region. This will have a positive feed-back effect on productivity, which will increase by a further 1% point, hence the total 5% effect mentioned above.

This discussion raises a second issue about how a policy-maker can increase the size of a region. The comparative static exercise we have just described corresponds to the compulsory displacement of populations. Clearly, that is not very realistic. The policy-maker must use endogenous market incentives for moving to larger cities, but little is known about these incentives. Theory would suggest that to increase the density one should alleviate some of the dispersion forces. Clearly, as discussed in section 1.3 local public provision, production subsidy or infrastructure investment have an impact on density. Typically, an increased supply of local transport infrastructure, schools, health services, etc. would alleviate dispersion forced and therefore increase density.

The land/housing market, as mentioned in Pfluger and Sudekum (2008) and in Puga (1999) also act as an important dispersion force. In those models, the housing and the land markets are fully competitive. This means that land and housing prices reflect their scarcity. Urban planning and lots of regulations are making those prices less informative about the real scarcity of the commodity. Reforms on that market may generate significant increase in density.

The real question is why should we try to influence its density? We know that there is a productivity gain that has to be traded off with the increased agglomeration cost linked to the increased density. The welfare discussion from section 1.2 shows, however, that for some of the parameter values, if there is an improvement to be made it is in having less agglomeration because of the efficiency and equity trade-off. Again, to assess overall welfare effects beyond productivity gains, one has to know where on the bell-shaped curve the economy is.

5.3 Implications for State aid

We saw in section 1.3 that infrastructure policies, local public good provision and tax incentives are likely to influence the agglomeration process. We also discussed the possibility that housing market regulation may affect the economic geography. The important question is whether local governments would be able to decide efficiently about those policies. A priori, the answer is negative as those policies, via their influence on agglomeration, generate inter-regional spillovers. Therefore, the non-cooperative equilibrium of the

game played by the different regions is likely to lead to an inefficient allocation of resources. Nevertheless, as the world considered here is of a second-best type, it is not obvious that the distortions introduced by the non-cooperative setting of local policies do not counteract other distortions.

In classic tax competition models, it is difficult to build the case for welfare improving state aids, for it rests on quite restrictive assumptions. One must be willing to assume away strategic interactions between governments, or consider that all politicians are self-motivated and rent maximisers. The assumptions behind the classic settings (constant returns to scale, perfectly competitive markets, costless trade) are, however, quite restrictive in themselves. Once one abandons them in favour of more realistic setups, there is scope for efficiency-enhancing fiscal competition – this simply stems from the usual second-best proviso that introducing further distortions in a distorted world may actually be desirable. Indeed, in imperfectly competitive settings like those we have considered in this survey, fiscal competition may have a corrective role. This happens if (i) the countries may design and commit to a tax schedule which is contingent on the full set of admissible firms strategies, (ii) firm subsidies are used to decrease the inefficiently high share of firms hosted by the core region, or (iii) to overcome inefficient locational lock-in. Put simply, when the subsidies are aimed at attracting mobile firms, they are bound to allow the firms to internalise the externalities that their mobility imposes on the remaining economic agents. This is no longer the case when a state grant helps to avoid the exit of the most inefficient firms following market integration. In such a case, the subsidies simply distort an efficient market mechanism. Hence, the presence of imperfect competition is not a sufficient condition for efficient tax competition.

Although both the European Union and the OECD have reacted against discriminatory taxation practices targeted at multinational firms, there is no consensus in the literature regarding the harmfulness of such practices. Discrimination may be desirable when firm mobility is endogenous. It also obtains when tax bases have different elasticities but none of them displays "home attachment", which potentially gives one of the countries enhanced market power over them. The desirability of restricting preferential regimes is ultimately an empirical question about what exactly differentiates tax bases. In any case, the policy implications one can draw from the literature are limited by the fact that it mostly relies on symmetrical country settings.

The regional competition to attract firms with productive public infrastructure seems to be wasteful in most contexts. Infrastructure provision is increased over and above the efficient level in terms of enhanced firm productivity, leading to a total provision cost which outweighs productivity gains. Again, one should be careful about devising policy recommendations based on the symmetric country framework. There are two exceptions to this insight. Firstly, when the economy is in a core-periphery equilibrium: in this case, the periphery actually under-provides infrastructure since it free-rides on the core's infrastructure, of which it can enjoy the benefits through lower import prices. This mechanism arises irrespective of whether agglomeration economies stem from mobility of the production factor or from the clustering of intermediate good suppliers. Secondly, allowing public infrastructure to be horizontally (i.e., different types of infrastructure as, for instance, the legal system, the transportation network, ...) or vertically (i.e., same type of infrastructure but differing in quality) differentiated may be used by the regions as a strategic device to soften tax competition, eventually promoting fully efficient tax setting and factor allocation in some cases. In this context, infrastructure serves the purpose of allowing the regions to capture a greater fiscal surplus from mobile production factors. Empirically, it seems to be the case that public inputs are not good mechanisms to attract capital, at least when accompanied by the corresponding,

budget-balanced, capital tax increase.

There are many ways in which countries can manipulate their policy choices as a means to soften tax competition. This survey identifies three of them: public input differentiation, public input under-provision, and discrimination of different tax bases. Policy-wise, this insight advises for caution when regulating other policy instruments, for one may unintentionally worsen the fiscal revenue dissipation of the tax competition equilibrium.

Finally, the existence of appropriately defined fiscal equalisation grants seems to mitigate the inefficiency of tax competition in a variety of settings, even when this latter is efficiency enhancing upon the no-tax equilibrium. The optimistic results are tilted in favour of tax base, as opposed to tax revenue, equalisation. While such grants already exist in many federal countries, there is scope for improving them along the lines identified in the literature. More importantly, the implementation of EU-wide fiscal equalisation schemes is bound to generate non-negligible efficiency gains.

There is also an empirical question. What is the real impact of the public policies aiming at attracting firms? The empirical literature here is not really optimistic. There is no work showing that infrastructure policies have a positive impact on location FDI. In the survey we acknowledge that this may be due to fact that no distinction is done between intra-national infrastructure and international ones. As theoretical predictions are different, it is not surprising that empirical studies are not conclusive. The evaluations of the impact of taxation are also plagued with the generic econometric problems linked to policy estimations. Two studies really try address those problems, Rathelot and Sillard (2008) and Duranton et al. (2011). They conclude that the impact of local taxes are quantitatively small.

More can be said about the impact of grants delivered directly to firms. Devereux et al. (2007) shows that the average impact of grants offered to firms is small but that the effect is way more important when some industrial agglomeration effect are at work. This means that subsidies are most effective to affect the decision of firms that would enjoy collocation benefits. Lastly, European funds are shown not to have an important average impact on the convergence between regions. However, interestingly, Becker et al. (2010) show that for regions that are at the limit of being eligible, the impact of the funds is significant. Rough calculations show that it gives a 20% return to the investment. This means that European funds if targeted to those regions that are either just eligible or just not eligible would give a higher return than the usual estimates.

Other studies may be used to assess the impact of state aids on firm locations. Criscuolo et al. (2012) study the impact of a UK manufacturing job support on productivity and employment, they show that the program has a positive effect on employment and investment but not on productivity. The absence of impact on productivity points to a possible weakening effect of the policy due to selection: The program maintains alive less productive firms. Another important result is that the program has an effect on firms that is differentiated according to their size. Only small firms benefit from the program. Mayer et al. (2011) also find that the impact of the urban French assistance programme is larger for small firms. Therefore restricting state aid to small firms seems to be a way to improve the efficiency of the policy since larger firms seems to be much less mobile and influenced by the programme. A possible explanation for that, which we discussed above, can be that more efficient and larger firms, do benefit more from agglomeration effects, as shown by Combes et al. (forthcoming). Therefore their cost to relocate in deprived, and generally smaller, areas is larger.

It is also interesting to point that most of the effect in the UK comes from creation of economic activity

rather than diversion from neighbouring regions. This means that those policies do not generate negative externalities on other regions. However, this is not what Mayer et al. (2011) find for France, where most of the effect comes from relocation at a very small geographical scale.

Importantly, the two studies on the French programme illustrate the possible presence of complementarities between the policy and other local characteristics of targeted areas, as Devereux et al. (2007) show for the UK for agglomeration effects. Mayer et al. (2011) show that the impact of the policy is stronger for targeted areas that are initially less distressed, usually the larger ones, and for sectors in which relocation costs are lower. Briant et al. (2011) show that a positive impact of the policy is obtained only when spatial isolation, which accounts for urban severance and transport access, is not too important.

The policy implications of such conclusions is interesting. Would the EU be willing to redistribute to ailing regions via state aids, those policies have to be targeted to particular regions and particular firms. Devereux et al. (2007) hint at targeting regions where some agglomeration already takes place, Becker et al. (2010) to regions that are at the limit of being eligible, Mayer et al. (2011) to areas that are initially less distressed and to sectors in which relocation costs are lower, Briant et al. (2011) to regions that have a good market access. All this is fully consistent with the literature on fiscal competition and economic geography. Indeed, it shows that when all the activity is agglomerated out of a region, firms enjoy an agglomeration rent in the other location and therefore the subsidies needed to relocate industry there is large. Subsidies may also be targeted to particular type of firms. Both Criscuolo et al. (2012) and Mayer et al. (2011) show that small firms are more responsive to subsidies. This is again in coherence with the literature showing that large firms can benefit more from economic density making them less willing to move to less dense regions.

Bibliography

Accetturo, A. and de Blasio, G. (2012). Policies for local development: An evaluation of Italy's "patti territoriali", *Regional Science and Urban Economics* **42**(1-2): 15–26.

Acs, Z., Audretsch, D. and Feldman, M. (1992). Real effects of academic research: Comment, *American Economic Review* **82**(1): 363–367.

Acs, Z., Audretsch, D. and Feldman, M. (1994). R&D spillovers and recipient firm size, *Review of Economics and Statistics* **100**(1): 15–29.

Adams, J. (2002). Comparative localization of academic and industrial spillovers, *Journal of Economic Geography* **2**(3): 253–278.

Adams, J. and Jaffe, A. (1996). Bounding the effects of R&D: An investigation using matched establishment-firm data, *Rand Journal of Economics* **27**(4): 700–721.

Anselin, L., Varga, A. and Acs, Z. (1997). Local geographic spillovers between university research and high technology innovations, *Journal of Urban Economics* **42**(3): 422–448.

Arauzo-Carod, J. and Viladecans-Marsal, E. (2009). Industrial location at the intra-metropolitan level: The role of agglomeration economies, *Regional Studies* **43**(4): 545–558.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations, *Review of Economic Studies* **58**(2): 277–297.

Arzaghi, M. and Henderson, J. V. (2008). Networking off Madison Avenue, *Review of Economic Studies* **75**(4): 1011–1038.

Au, C.-C. and Henderson, V. (2006). Are chinese cities too small?, *Review of Economic Studies* **73**(3): 549–576.

Audretsch, D. and Feldman, M. (1996). R&D spillovers and the geography of innovation and production, *American Economic Review* **86**(3): 630–640.

Bacolod, M., Blum, B. S. and Strange, W. C. (2009a). Skills in the city, *Journal of Urban Economics* **65**(2): 136–153.

Bacolod, M., Blum, B. S. and Strange, W. C. (2009b). Urban interactions: soft skills versus specialization, *Journal of Economic Geography* **9**(2): 227–262.

Baldwin, R. E., Forslid, R., Martin, P., Ottaviano, G. I. P. and Robert-Nicoud, F. (2004). *Economic Geography and Public Policy*, Princeton University Press, New Jersey.

Baldwin, R. E. and Krugman, P. (2004). Agglomeration, integration and tax harmonisation, *European Economic Review* **48**(1): 1–23.

Barrios, S., Görg, H. and Strobl, E. (2006). Multinationals' location choice, agglomeration economies, and public incentives, *International Regional Science Review* **29**(1): 81–107.

Bartelsman, E. J. and Beetsma, R. (2003). Why pay more? Corporate tax avoidance through transfer pricing in OECD countries, *Journal of Public Economics* **87**(9-10): 2225–2252.

Basile, R. (2004). Acquisition versus greenfield investment: The location of foreign manufacturers in Italy, *Regional Science and Urban Economics* **34**(1): 3–25.

Basile, R., Castellani, D. and Zanfei, A. (2008). Location choices of multinational firms in europe: The role of EU cohesion policy, *Journal of International Economics* **74**(2): 328–340.

Basile, R., Castellani, D. and Zanfei, A. (2009). National boundaries and the location of multinational firms in europe, *Papers in Regional Science* **88**(4): 733–748.

Becker, S. O., Egger, P. H. and von Ehrlich, M. (2010). Going nuts: The effect of EU structural funds on regional performance, *Journal of Public Economics* **94**(9-10): 578–590.

Behrens, K., Gaigné, C., Ottaviano, G. I. and Thisse, J.-F. (2007). Countries, regions and trade: On the welfare impacts of economic integration, *European Economic Review* **51**(5): 1277–1301.

Benassy-quere, A., Gobalraja, N. and Trannoy, A. (2007). Tax and public input competition, *Economic Policy,* **22**(50): 491–507.

Besley, T. and Seabright, P. (1999). The effects and policy implications of state aids to industry: An economic analysis, *Economic Policy,* **14**(28).

Besley, T. and Smart, M. (2007). Fiscal restraints and voter welfare, *Journal of Public Economics* **91**(3-4): 755–773.

Blien, U. and Suedekum, J. (2005). Local economic structure and industry development in germany, 1993-2001, *Economics Bulletin* **17**: 1–8.

Blien, U., Suedekum, J. and Wolf, K. (2006). Productivity and the density of economic activity, *Labour Economics* **13**(4): 445–458.

Boadway, R. W. and Flatters, F. R. (1982). Efficiency and equalization payments in a federal system of government: A synthesis and extension of recent results, *Canadian Journal of Economics* **15**(4): 613–33.

Bode, E. (2004). The spatial pattern of localized R&D spillovers: an empirical investigation for Germany, *Journal of Economic Geography* **4**(1): 43–64.

Borck, R., Caliendo, M. and Steiner, V. (2007). Fiscal competition and the composition of public spending: Theory and evidence, *FinanzArchiv: Public Finance Analysis* **63**(2): 264–277.

Borck, R., Koh, H.-J. and Pflüger, M. (2009). Inefficient lock-in and subsidy competition, *Working Paper 2592*, CESifo.

Bottazzi, L. and Peri, G. (2003). Innovation and spillovers: Evidence from european regions, *European Economic Review* **47**: 687–710.

Boutin, X., Cestone, G., Fumagalli, C., Pica, G. and Serrano-Velarde, N. (2009). The deep pocket effect of internal capital markets, *Discussion papers*, CEPR.

Brakman, S., Garretsen, H. and Schramm, M. (2004). The spatial distribution of wages: Estimating the helpman-hanson model for Germany, *Journal of Regional Science* **44**(3): 437–466.

Breinlich, H. (2006). The spatial income structure in the european union: What role for economic geography?, *Journal of Economic Geography* **6**(5): 593–617.

Brennan, G. and Buchanan, J. M. (1980). *The Power to Tax*, Cambridge University Press.

Briant, A., Lafourcade, M. and Schmutz, B. (2011). Urban geography and the effectiveness of french enterprise zones. Processed, Paris School of Economics.

Brülhart, M. and Mathys, N. A. (2008). Sectoral agglomeration economies in a panel of European regions, *Regional Science and Urban Economics* **38**(4): 348–362.

Buch, C. M., Kleiner, J., Lipponer, A. and Toubal, F. (2006). Determinants and effects of foreign direct investment: evidence from german firm-level data, *Economic Policy* **20**: 52 – 110.

Bucovetsky, S. (1991). Asymmetric tax competition, *Journal of Urban Economics* **30**(2): 167–181.

Bucovetsky, S. (2005). Public input competition, *Journal of Public Economics* **89**(9-10): 1763–1787.

Bucovetsky, S. and Haufler, A. (2008). Tax competition when firms choose their organizational form: Should tax loopholes for multinationals be closed?, *Journal of International Economics* **74**(1): 188–201.

Bucovetsky, S. and Smart, M. (2006). The efficiency consequences of local revenue equalization: Tax competition and tax distortions, *Journal of Public Economic Theory* **8**(1): 119–144.

Bucovetsky, S. and Wilson, J. D. (1991). Tax competition with two tax instruments, *Regional Science and Urban Economics* **21**(3): 333–350.

Buettner, T. (2006). The incentive effect of fiscal equalization transfers on tax policy, *Journal of Public Economics* **90**(3): 477–497.

Carlino, G. A., Chatterjee, S. and Hunt, R. M. (2007). Urban density and the rate of invention, *Journal of Urban Economics* **61**(3): 389–419.

Carlton, D. (1983). The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables, *Review of Economics and Statistics* **65**: 440–449.

Charlot, S. and Duranton, G. (2004). Communication externalities in cities, *Journal of Urban Economics* **56**(3): 581–613.

Charlot, S., Gaigné, C., Robert-Nicoud, F. and Thisse, J.-F. (2006). Agglomeration and welfare: The core-periphery model in the light of Bentham, Kaldor, and Rawls, *Journal of Public Economics* **90**(1-2): 325–347.

Ciccone, A. (2002). Agglomeration effects in Europe, *European Economic Review* **46**(2): 213–227.

Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity, *American Economic Review* **86**(1): 54–70.

Ciccone, A. and Peri, G. (2006). Identifying human capital externalities: Theory with an application to US cities, *Review of Economic Studies* **73**: 381–412.

Cieślik, A. (2005). Regional characteristics and the location of foreign firms within poland, *Applied Economics* **37**(8): 863–874.

Cingano, F. and Schivardi, F. (2004). Identifying the sources of local productivity growth, *Journal of the European Economic Association* **2**(4): 720–742.

Clausing, K. A. (2003). Tax-motivated transfer pricing and US intrafirm trade prices, *Journal of Public Economics* **87**(9-10): 2207–2223.

Combes, P.-P. (2000). Economic structure and local growth: France, 1984–1993, *Journal of Urban Economics* **47**(3): 329–355.

Combes, P.-P., Duranton, G. and Gobillon, L. (2008). Spatial wage disparities: Sorting matters!, *Journal of Urban Economics* **63**: 723–742.

Combes, P.-P., Duranton, G. and Gobillon, L. (2011). The identification of agglomeration economies, *Journal of Economic Geography* **11**: 253–266.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. and Roux, S. (forthcoming). The productivity advantages of large markets: Distinguishing agglomeration from firm selection, *Econometrica* .

Combes, P.-P., Duranton, G., Gobillon, L. and Roux, S. (2010). Estimating agglomeration effects with history, geology, and worker fixed-effects, *in* E. L. Glaeser (ed.), *Agglomeration Economics*, Chicago University Press, Chicago, IL, pp. 15–65.

Combes, P.-P., Magnac, T. and Robin, J.-M. (2004). The dynamics of local employment in France, *Journal of Urban Economics* **56**(2): 217–243.

Combes, P.-P., Mayer, T. and Thisse, J.-F. (2008). *Economic Geography: The integration of Regions and Nations*, Princeton University Press, New Jersey.

Costa-Campi, M., Segarra-Blasco, A. and Viladecans-Marsal, E. (2004). The location of new firms and the life cycle of industries, *Small Business Economics* **22**(3-4): 265–281.

Cremer, H. and Gahvari, F. (2000). Tax evasion, fiscal competition and economic integration, *European Economic Review* **44**(9): 1633–1657.

Criscuolo, C., Martin, R., Overman, H. and Reenen, J. V. (2012). The causal effect of an industrial policy, *Working Paper 17842*, National Bureau of Economic Research.

Crozet, M., Mayer, T. and Mucchielli, J.-L. (2004). How do firms agglomerate? A study of FDI in France, *Regional Science and Urban Economics* **34**(1): 27–54.

Dall'erba, S. (2005). Distribution of regional income and regional funds in Europe 1989-1999: An exploratory spatial data analysis, *The Annals of Regional Science* **39**: 121–148.

Dall'erba, S. and Le Gallo, J. (2008). Regional convergence and the impact of European structural funds over 1989-1999: A spatial econometric analysis, *Papers in Regional Science* **87**(2): 219–244.

Desmet, K. and Fafchamps, M. (2005). Changes in the spatial concentration of employment across US counties: a sectoral analysis 1972-2000, *Journal of Economic Geography* **5**(3): 261–284.

Devereux, M. P., Griffith, R. and Simpson, H. (2007). Firm location decisions, regional grants and agglomeration externalities, *Journal of Public Economics* **91**(3-4): 413–435.

Devereux, M. P., Lockwood, B. and Redoano, M. (2008). Do countries compete over corporate tax rates?, *Journal of Public Economics* **92**(5-6): 1210–1235.

Disdier, A.-C. and Mayer, T. (2004). How different is Eastern Europe? structure and determinants of location choices by French firms in Eastern and Western europe, *Journal of Comparative Economics* **32**(1): 280 – 296.

Duflo, E., Glennerster, R. and Kremer, M. (2007). Using randomization in development economics research: A toolkit, *in* T. P. Schultz and J. A. Strauss (eds), *Handbook of Development Economics*, Vol. 4, Elsevier, pp. 3895–3962.

Duranton, G., Gobillon, L. and Overman, H. G. (2011). Assessing the effects of local taxation using microgeographic data, *The Economic Journal* **121**(555): 1017–1046.

Duranton, G. and Monastiriotis, V. (2002). Mind the gaps: The evolution of regional earnings inequalities in the UK 1982–1997, *Journal of Regional Science* **42**(2): 219–256.

Duranton, G. and Turner, M. (2011). Urban growth and transportation. Mimeographed, University of Toronto.

Edwards, J. and Keen, M. (1996). Tax competition and Leviathan, *European Economic Review* **40**(1): 113–134.

Egger, H. and Falkinger, J. (2006). The role of public infrastructure and subsidies for firm location and international outsourcing, *European Economic Review* **50**(8): 1993–2015.

Egger, P., Koethenbuerger, M. and Smart, M. (2010). Do fiscal transfers alleviate business tax competition? Evidence from Germany, *Journal of Public Economics* **94**(3-4): 235–246.

Esposti, R. and Bussoletti, S. (2008). Impact of objective 1 funds on regional growth convergence in the European union: A panel-data approach, *Regional Studies* **42**(2): 159–173.

European Parliament (2007). Regional disparities and cohesion: What strategies for the future, *Report for the DG for internal policies of the Union*.

EuropeanCommunities (1998). Conclusions of the ECOFIN council meeting on 1 december 1997 concerning taxation policy (including Code of Conduct for Business Taxation), *Official Journal of the European Communities* **98/C**(2/01).

Feldman, M. (1994). *The Geography of Innovation*, Kluwer Academic Publishers, Boston.

Feldman, M. and Audretsch, D. (1999). Innovation in cities: Science-based diversity, specialization and localized competition, *European Economic Review* **43**(2): 409–429.

Fenge, R., von Ehrlich, M. and Wrede, M. (2009). Public input competition and agglomeration, *Regional Science and Urban Economics* **39**(5): 621–631.

Figueiredo, O., aes, P. G. and Woodward, D. (2002). Home-field advantage: location decisions of portuguese entrepreneurs, *Journal of Urban Economics* **52**(2): 341–361.

Fingleton, B. (2003). Increasing returns: evidence from local wage rates in great britain, *Oxford Economic Papers* **55**(4): 716–739.

Fischer, M., Scherngell, T. and Jansenberger, E. (2006). The geography of knowledge spillovers between high technology firms in europe: Evidence from a spatial interaction modelling perspective, *Geographical Analysis* **38**: 288–309.

Fischer, M., Scherngell, T. and Jansenberger, E. (2009a). Geographic localisation of knowledge spillovers: evidence from high-tech patent citations in europe, *The Annals of Regional Science* **43**: 839–858.

Fischer, M., Scherngell, T. and Jansenberger, E. (2009b). Patents, patent citations and the geography of knowledge spillovers in europe, *in* C. Karlsson, A. Andersson, P. Cheshire and R. Stough (eds), *New Directions in Regional Economic Development*, Springer Verlag, Berlin, pp. 331–345.

Forslid, R., Haaland, J. I. and Knarvik, K. H. M. (2002). A U-shaped Europe? a simulation study of industrial location, *Journal of International Economics* **57**(2): 273–297.

Forslid, R. and Ottaviano, G. I. (2003). An analytically solvable core-periphery model, *Journal of Economic Geography* **3**(3): 229–240.

Foster, N. and Stehrer, R. (2009). Sectoral productivity, density and agglomeration in the Wider Europe, *Spatial Economic Analysis* **4**(4): 427–446.

Fuchs, M. (2011). The determinants of local employment dynamics in Western Germany, *Empirical Economics* **40**(1): 177–203.

Fujita, M., Krugman, P. R. and Venables, A. J. (1999). *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press, Cambridge, MA.

Gabe, T. M. and Bell, K. P. (2004). Tradeoffs between local taxes and government spending as determinants of business location, *Journal of Regional Science* **44**(1): 21–41.

Gaigné, C. and Riou, S. (2007). Globalization, asymmetric tax competition, and fiscal equalization, *Journal of Public Economic Theory* **9**(5): 901–925.

Glaeser, E. L., Kallal, H., Scheinkman, J. A. and Schleifer, A. (1992). Growth in cities, *Journal of Political Economy* **100**(6): 1126–1152.

Glaeser, E. L. and Maré, D. C. (2001). Cities and skills, *Journal of Labor Economics* **19**(2): 316–342.

Glaeser, E. L., Rosenthal, S. S. and Strange, W. C. (2010). Urban economics and entrepreneurship, *Journal of Urban Economics* **67**(1): 1–14.

Graham, D. J. (2007). Variable returns to agglomeration and the effect of road traffic congestion, *Journal of Urban Economics* **62**(1): 103–120.

Graham, D. J., Melo, P. S., Jiwattanakulpaisarn, P. and Noland, R. B. (2010). Testing for causality between productivity and agglomeration economies, *Journal of Regional Science* **50**(5): 935–951.

Greunz, L. (2003). Geographically and technologically mediated knowledge spillovers between European regions, *The Annals of Regional Science* **37**: 657–680.

Guimaraes, P., Figueiredo, O. and Woodward, D. (2000). Agglomeration and the location of foreign direct investment in portugal, *Journal of Urban Economics* **47**(1): 115–135.

Hamilton, J., Thisse, J.-F. and Zenou, Y. (2000). Wage competition with heterogeneous workers and firms, *Journal of Labor Economics* **18**(3): 453–72.

Hanson, G. H. (1996). Economic integration, intraindustry trade, and frontier regions, *European Economic Review* **40**(3–5): 941–949.

Hanson, G. H. (1997). Increasing returns, trade, and the regional structure of wages, *Economic Journal* **107**(440): 113–133.

Hanson, G. H. (2005). Market potential, increasing returns, and geographic concentration, *Journal of International Economics* **67**(1): 1–24.

Hanson, N. (1990). Do producer services induce regional development?, *Journal of Regional Science* **30**: 465–478.

Harris, C. (1954). The market as a factor in the localization of industry in the United States, *Annals of the Association of American Geographers* **44**(4): 315–348.

Haupt, A. and Peters, W. (2005). Restricting preferential tax regimes to avoid harmful tax competition, *Regional Science and Urban Economics* **35**(5): 493–507.

Head, C., Ries, J. C. and Swenson, D. L. (1999). Attracting foreign manufacturing: Investment promotion and agglomeration, *Regional Science and Urban Economics* **29**(2): 197–218.

Head, K. and Mayer, T. (2004). Market potential and the location of japanese investment in the European union, *Review of Economics and Statistics* **86**(4): 959–972.

Head, K. and Mayer, T. (2006). Regional wage and employment responses to market potential in the eu, *Regional Science and Urban Economics* **36**(5): 573–595.

Heckman, J. J., Lalonde, R. J. and Smith, J. A. (1999). The economics and econometrics of active labor market programs, *in* O. C. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3, Part 1, Elsevier, pp. 1865–2097.

Hellerstein, J. K., Neumark, D. and Troske, K. R. (1999). Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations, *Journal of Labour Economics* **17**(3): 409–446.

Helpman, E. (1998). The size of regions, *in* D. Pines, E. Sadka and I. Zilcha (eds), *Topics in Public Economics. Theoretical and Applied Analysis*, Cambridge University Press, New York, NY, pp. 33–54.

Henderson, J. (1997). Externalities and industrial development, *Journal of Urban Economics* **42**(3): 449–470.

Henderson, J. (2003). Marshall's economies, *Journal of Urban Economics* **53**(1): 1–28.

Henderson, J., Kuncoro, A. and Turner, M. (1995). Industrial development in cities, *Journal of Political Economy* **103**(5): 1067–1090.

Hilber, C. A. L. and Voicu, I. (2010). Agglomeration economies and the location of foreign direct investment: Empirical evidence from Romania, *Regional Studies* **44**(3): 355–371.

Hindriks, J. and Myles, G. D. (2003). Strategic inter-regional transfers, *Journal of Public Economic Theory* **5**(2): 229–248.

Hindriks, J., Peralta, S. and Weber, S. (2008). Competing in taxes and investment under fiscal equalization, *Journal of Public Economics* **92**(12): 2392–2402. New Directions in Fiscal Federalism.

Hines, J. R. (1997). Tax policy and the activities of multinational corporations, *in* Auerbach (ed.), *Fiscal Policy: Lessons from Economic Research*, MIT press, Cambridge.

Hines, J. R. (1999). Lessons from behavioral responses to international taxation, *National Tax Journal* **52**(2): 305–332.

Holl, A. (2004). Manufacturing location and impacts of road transport infrastructure: Empirical evidence from spain, *Regional Science and Urban Economics* **34**(3): 341–363.

Holl, A. (forthcoming). Market potential and firm-level productivity in Spain, *Journal of Economic Geography* .

Hong, Q. and Smart, M. (2010). In praise of tax havens: International tax planning and foreign direct investment, *European Economic Review* **54**(1): 82–95.

Irmen, A. and Thisse, J.-F. (1998). Competition in multi-characteristics spaces: Hotelling was almost right, *Journal of Economic Theory* **78**(1): 76–102.

Jacobs, J. (1969). *The Economy of Cities*, Random House, New York.

Jaffe, A. B. (1989). Real effects of academic research, *American Economic Review* **79**(5): 957–970.

Jaffe, A. B. and Trajtenberg, M. (2002). *Patents, Citations and Innovations: A Window on the Knowledge Economy*, MIT Press, Cambridge, Mass.

Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* **108**(3): 577–598.

James, D. and Gordon, M. (1994). Strategic capital tax competition: A pecuniary externality and a corrective device, *Journal of Urban Economics* **36**(1): 66–78.

Janeba, E. and Peters, W. (1999). Tax evasion, tax competition and the gains from nondiscrimination: The case of interest taxation in Europe, *Economic Journal* **109**(452): 93–101.

Janeba, E. and Smart, M. (2003). Is targeted tax competition less harmful than its remedies?, *International Tax and Public Finance* **10**(3): 259–80.

Jurajda, S. and Terrell, K. (2009). Growth in cities, *Regional unemployment and human capital in transition economies* **17**(2): 241–274.

Justman, M., Thisse, J.-F. and van Ypersele, T. (2005). Fiscal competition and regional differentiation, *Regional Science and Urban Economics* **35**(6): 848–861.

Justman, M., Thisse, J.-F. and vanYpersele, T. (2002). Taking the bite out of fiscal competition, *Journal of Urban Economics* **52**(2): 294–315.

Keen, M. (1993). The welfare economics of tax co-ordination in the European Community : A survey, *Fiscal Studies* **14**(2): 15–36.

Keen, M. (2001). The effects and policy implications of state aids to industry: An economic analysis, *National Tax Journal,* **54**.

Keen, M. and Marchand, M. (1997). Fiscal competition and the pattern of public spending, *Journal of Public Economics* **66**(1): 33–53.

Kind, H. J., Midelfart-Knarvik, K.-H. and Schjelderup, G. (2000). Competing for capital in a 'lumpy' world, *Journal of Public Economics* **78**(3): 253–274.

Kline, P. (2010). Place based policies, heterogeneity, and agglomeration, *American Economic Review* **100**(2): 383–87.

Kothenburger, M. (2002). Tax competition and fiscal equalization, *International Tax and Public Finance* **9**(4): 391–408.

Krogstrup, S. (2003). Are capital taxes racing to the bottom in the European Union?, *Working Papers 01-2003*, IHEID.

Krugman, P. (1991a). History versus expectations, *The Quarterly Journal of Economics* **106**(2): 651–67.

Krugman, P. (1992). *Geography and Trade*, The MIT Press.

Krugman, P. R. (1991b). Increasing returns and economic geography, *Journal of Political Economy* **99**(3): 484–499.

Krugman, P. R. and Venables, A. J. (1995). Globalization and the inequality of nations, *Quarterly Journal of Economics* **110**(4): 857–880.

LeSage, J. P., Fischer, M. M. and Scherngell, T. (2007). Knowledge spillovers across europe: Evidence from a poisson spatial interaction model with spatial effects, *Papers in Regional Science* **86**(3): 393–421.

Liesegang, C. and Runkel, M. (2009). Corporate income taxation of multinationals and fiscal equalization, *Working Paper Series 2747*, CESifo.

Ludema, R. D. and Wooton, I. (2000). Economic geography and the fiscal effects of regional integration, *Journal of International Economics* **52**(2): 331–357.

Maggioni, M. A., Nosvelli, M. and Uberti, T. E. (2007). Space versus networks in the geography of innovation: A European analysis, *Papers in Regional Science* **86**(3): 471–493.

Mameli, F., Faggian, A. and McCann, P. (2008). Employment growth in italian local labour systems: Issues of model specification and sectoral aggregation, *Spatial Economic Analysis* **3**(3): 343–360.

Mansori, K. S. and Weichenrieder, A. J. (2001). Tax competition and transfer pricing disputes, *FinanzArchiv: Public Finance Analysis* **58**(1): 1–11.

Marshall, A. (1890). *Principles of Economics*, Macmillan, London.

Martin, P. (1999). Public policies, regional inequalities and growth, *Journal of Public Economics* **73**(1): 85–105.

Martin, P., Mayer, T. and Mayneris, F. (2011a). Public support to clusters: A firm level study of French local productive systems, *Regional Science and Urban Economics* **41**(2): 108–123.

Martin, P., Mayer, T. and Mayneris, F. (2011b). Spatial concentration and plant-level productivity in france, *Journal of Urban Economics* **69**(2): 182–195.

Martin, P. and Ottaviano, G. I. (1999). Growing locations: Industry location in a model of endogenous growth, *European Economic Review* **43**(2): 281–302.

Martin, P. and Ottaviano, G. I. (2001). Growth and agglomeration, *International Economic Review* **42**(4): 947–68.

Martin, P. and Rogers, C. A. (1995). Industrial location and public infrastructure, *Journal of International Economics* **39**(3–4): 335–351.

Martin, S. and Valbonesi, P. (2008). Equilibrium state aid in integrating markets, *The B.E. Journal of Economic Analysis & Policy* **8**(1).

Maurseth, P. B. and Verspagen, B. (2002). Knowledge spillovers in europe: A patent citations analysis, *Scandinavian Journal of Economics* **104**(4): 531–545.

Mayer, T., Maynéris, F. and Py, L. (2011). The impact of urban enterprise zones on establishment location decisions - evidence from french ZFUs. Processed, Sciences Po Paris.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka (ed.), *Frontier in Econometrics*, Academic Press, New York, pp. 105–142.

McFadden, D. (1980). Econometric models for probabilistic choice among products, *Journal of Business* **53**(3): S13–S29.

Melo, P. C., Graham, D. J. and Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies, *Regional Science and Urban Economics* **39**(3): 332–342.

Micucci, G. and Giacinto, V. (2009). The producer service sector in Italy: Long-term growth and its local determinants, *Spatial Economic Analysis* **4**(4): 391–425.

Mion, G. (2004). Spatial externalities and empirical analysis: The case of Italy, *Journal of Urban Economics* **56**(1): 97–118.

Mion, G. and Naticchioni, P. (2009). The spatial sorting and matching of skills and firms, *Canadian Journal of Economics* **42**: 28–55.

Mohl, P. and Hagen, T. (2010). Do EU structural funds promote regional growth? new evidence from various panel data approaches, *Regional Science and Urban Economics* **40**(5): 353–365.

Monfort, P. and Nicolini, R. (2000). Regional convergence and international integration, *Journal of Urban Economics* **48**(2): 286–306.

Monfort, P. and van Ypersele, T. (2003). Integration, regional agglomeration and international trade, *Discussion Papers 3752*, CEPR.

Moretti, E. (2004a). Human capital externalities in cities, *in* V. Henderson and J.-F. Thisse (eds), *Handbook of Regional and Urban Economics*, Vol. 4, North-Holland, Amsterdam, pp. 2243–2291.

Moretti, E. (2004b). Workers' education, spillovers, and productivity: Evidence from plant-level production functions, *American Economic Review* **94**(3): 656–690.

Murata, Y. (2003). Product diversity, taste heterogeneity, and geographic distribution of economic activities:: market vs. non-market interactions, *Journal of Urban Economics* **53**(1): 126–144.

Ottaviano, G. I. and Thisse, J.-F. (2002). Integration, agglomeration and the political economics of factor mobility, *Journal of Public Economics* **83**(3): 429–456.

Ottaviano, G. I. and van Ypersele, T. (2005). Market size and tax competition, *Journal of International Economics* **67**(1): 25–46.

Pakes, A. and Griliches, Z. (1984). Patents and r&d at the firm level: A first look, *in* Z. Griliches (ed.), *R&D, Patents, and Productivity*, University of Chicago Press, Chicago.

Parent, O. and Riou, S. (2005). Bayesian analysis of knowledge spillovers in European regions, *Journal of Regional Science* **45**(4): 747–775.

Peralta, S. and van Ypersele, T. (2005). Factor endowments and welfare levels in an asymmetric tax competition game, *Journal of Urban Economics* **57**(2): 258–274.

Peralta, S. and van Ypersele, T. (2006). Coordination of capital taxation among asymmetric countries, *Regional Science and Urban Economics* **36**(6): 708–726.

Peralta, S., Wauthy, X. and van Ypersele, T. (2006). Should regions control international profit shifting?, *Journal of International Economics* **68**(24).

Pfluger, M. and Sudekum, J. (2008). Integration, agglomeration and welfare, *Journal of Urban Economics* **63**(2): 544–566.

Picard, P. M. and Zeng, D.-Z. (2005). Agricultural sector and industrial agglomeration, *Journal of Development Economics* **77**: 75–106.

Porter, M. E. (1990). *The Competitive Advantage of Nations*, Free Press, New York.

Porter, M. E. (1998). Clusters and the new economics of competition, *Harvard Business Review* **76**(6): 77–91.

Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy, *Economic Development Quarterly* **14**(1): 15–34.

Puga, D. (1999). The rise and fall of regional inequalities, *European Economic Review* **43**(2): 303–334.

Puigcerver-Penalver, M. C. (2007). The impact of structural funds policy on European regions' growth. a theoretical and empirical approach, *The European Journal of Comparative Economics* **4**: 179208.

Pusterla, F. and Resmini, L. (2007). Where do foreign firms locate in transition countries? An empirical investigation, *The Annals of Regional Science* **41**: 835–856.

Raimondos-Moller, P. (2002). Transfer pricing rules and competing governments, *Oxford Economic Papers* **54**(2): 230–246.

Ramajo, J., Márquez, M. A., Hewings, G. J. and Salinas, M. M. (2008). Spatial heterogeneity and interregional spillovers in the European union: Do cohesion policies encourage convergence across regions?, *European Economic Review* **52**(3): 551–567.

Rathelot, R. and Sillard, P. (2008). The importance of local corporate taxes in business location decisions: Evidence from French micro data, *The Economic Journal* **118**(527): 499–514.

Redding, S. and Sturm, D. (2008). The costs of remoteness: Evidence from german division and reunification, *American Economic Review* **98**(5): 1766–1797.

Redding, S. and Venables, A. J. (2004). Economic geography and international inequality, *Journal of International Economics* **62**(1): 63–82.

Rice, P., Venables, A. J. and Patacchini, E. (2006). Spatial determinants of productivity: Analysis for the regions of great britain, *Regional Science and Urban Economics* **36**(6): 727–752.

Roback, J. (1982). Wages, rents and the quality of life, *Journal of Political Economy* **90**(6): 1257–1278.

Rodríguez-Pose, A. and Fratesi, U. (2004). Between development and social policies: The impact of European structural funds in objective 1 regions, *Regional Studies* **38**(1): 97–113.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, industrial agglomeration, and agglomeration, *Review of Economics and Statistics* **85**(2): 377–393.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies, *in* V. Henderson and J.-F. Thisse (eds), *Handbook of Regional and Urban Economics*, Vol. 4, North-Holland, Amsterdam, pp. 2119–2171.

Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers, *Journal of Urban Economics* **64**(2): 373–389.

Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Harvard University Press, Cambridge, MA.

Slemrod, J. and Wilson, J. D. (2009). Tax competition with parasitic tax havens, *Journal of Public Economics* **93**(11-12): 1261–1270.

Smart, M. (1998). Taxation and deadweight loss in a system of intergovernmental transfers, *Canadian Journal of Economics* **31**(1): 189–206.

Solé-Ollé, A. and Viladecans-Marsal, E. (2004). Central cities as engines of metropolitan area growth, *Journal of Regional Science* **44**(2): 321–350.

Spies, J. (2010). Network and border effects: Where do foreign multinationals locate in Germany?, *Regional Science and Urban Economics* **40**(1): 20–32.

Starrett, D. (1978). Market allocations of location choice in a model with free mobility, *Journal of Economic Theory* **17**: 21–37.

Tabuchi, T. and Thisse, J.-F. (2002). Taste heterogeneity, labor mobility and economic geography, *Journal of Development Economics* **69**(1): 155–177.

Thompson, P. (2006). Patent citations and the geography of knowledge spillovers: Evidence from inventor- and examiner-added citations, *Review of Economics and Statistics* **88**(2): 383–389.

Tiebout, C. M. (1956). A pure theory of local expenditures, *Journal of Political Economy* **64**(5): 416–424.

Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA.

Verspagen, B. and Schoenmakers, W. (2004). The spatial dimension of patenting by multinational firms in europe, *Journal of Economic Geography* **4**(1): 23–42.

Viladecans-Marsal, E. (2004). Agglomeration economies and industrial location: City-level evidence, *Journal of Economic Geography* **5**(4): 565–582.

Wildasin, D. E. (1988). Nash equilibria in models of fiscal competition, *Journal of Public Economics* **35**(2): 229–240.

Wildasin, D. E. (1989). Interjurisdictional capital mobility: Fiscal externality and a corrective subsidy, *Journal of Urban Economics* **25**(2): 193–212.

Wildasin, D. E. (1991). Some rudimetary 'duopolity' theory, *Regional Science and Urban Economics* **21**(3): 393–421.

Wilson, J. D. (1987). Trade, capital mobility and tax competition, *Journal of Political Economy* **95**(4): 835–56.

Wilson, J. D. (1991). Tax competition with interregional differences in factor endowments, *Regional Science and Urban Economics* **21**(3): 423–451.

Wilson, J. D. and Wildasin, D. (2004). Capital tax competition: bane or boon, *Journal of Public Economics* **88**(6): 1065–1091.

Wren, C. and Jones, J. (2009). Re-investment and the survival of foreign-owned plants, *Regional Science and Urban Economics* **39**(2): 214–223.

Wren, C. and Jones, J. (2011). Assessing the regional impact of grants on FDI location: Evidence from UK regional policy, 1985-2005, *Journal of Regional Science* **51**(3): 497–517.

Zodrow, G. R. and Mieszkowski, P. (1986). Pigou, Tiebout, property taxation, and the underprovision of local public goods, *Journal of Urban Economics* **19**(3): 356–370.

European Commission

The role and effectiveness of regional investment aid - The point of view of the academic literature

Luxembourg: Publications Office of the European Union

2013 – 132 pp. – 21.0 x 29.7 cm

ISBN 978-92-79-28198-3

doi: 10.2763/32058

Abstract

The purpose of this report is to study, from both the theoretical and the empirical point of view, the extent to which regional policy can reduce the disparities in economic activity levels that arise between regions belonging to an integrated trade area as the European Union.

KD-32-13-024-EN-N

10.2763/32058