HAL
open science

The empirics of economic geography: how to draw policy implications?

Pierre-Philippe Combes

To cite this version:

Pierre-Philippe Combes. The empirics of economic geography: how to draw policy implications?. Review of World Economics, 2011, 147 (3), pp.567-592. 10.1007/s10290-011-0092-z . hal-00676677

HAL Id: hal-00676677

https://hal.science/hal-00676677v1

Submitted on 6 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The empirics of economic geography: How to draw policy implications? *

Pierre-Philippe Combes†

21st February 2011

Abstract

Using both reduced-form and structural approaches, the spectrum of policy recommendations that can be drawn from empirical economic geography is pretty large. Reduced-form approaches allow the researchers to consider many variables that impact on regional disparities, as long as they are careful about interpretation and endogeneity issues. Structural approaches have the opposite advantages. Less issues can be simultaneously addressed, but one can be more precise in terms of which intuitions are considered and the underlying mechanisms and effects at work. Many regional policy issues remain unanswered, opening some interesting future lines of research.

Key words: agglomeration economies, regional policy, structural estimation, instrumental variables

JEL classification: R12, R23

*Paper prepared for the VII Annual Elsnit Conference hosted by the Kiel Institute for the World Economy, under the sponsorship and co-organisation of the Inter-American Development Bank (IDB), October 2009. Comments from Rolf Langhammer, Harmen Lehment, Elsnit conference participants and a referee were appreciated. The presentation of empirical approaches in economic geography and their limits (sections 2.1 and 3.1) borrows some elements from Combes, Mayer and Thisse, *Economic Geography, The Integration of regions and Nations*, Princeton University Press, 2008. Financial support from IDB and CNRS is gratefully acknowledged.

†GREQAM-University of Aix-Marseille. Also affiliated with the Centre for Economic Policy Research, International Trade and Regional Economics programme. ppcombes@univmed.fr. http://www.vcharite.univ-mrs.fr/pp/combes/.

1 Introduction

In their book, Fujita et al. (1999) called for the emergence of empirical contributions to assess the relevance of new economic geography models. Ten years later, quite a few papers have shown that such models are not rejected by the real data. Moreover, these papers emphasise the role of certain variables, such as the real market potential, that are extremely relevant empirically to explain regional development. Beyond such conclusions, we try here to assess the policy implications to be drawn from this empirical material, the limits to such exercises and the directions research could take to better enlighten the regional policy debate.

We start in Section 2 by describing an empirical literature whose roots are older than new economic geography models, but which has benefited from renewed interest with this emphasis on the need for empirical assessment of regional development models. Typically, this literature seeks to determine which characteristics of a region enhance its productivity or growth. We show that advances in econometric methods and the availability of individual data sets has allowed significant progress in the precision and robustness of conclusions, and therefore on the policy implications that can be drawn. We then review in Section 3 the more structural empirical literature, sticking closely to new economic geography models, and we illustrate how it can be used to derive relevant policy implications. In both cases, we conclude that empirical studies specifically designed to deal with policy issues are still the exception and we propose possible lines of research in such a direction.

2 Reduced-form approaches and the optimal characteristics of regions

In recent years, interest in the empirical evaluation of new economic geography models has in turn revived interest in the literature estimating the magnitude of agglomeration economies, which is older and with fairly different theoretical roots. To assess the empirical relevance of these reduced-form approaches, we first describe their theoretical background, define the empirical assumptions and how they are dealt with. Then we move to possible policy implications of the results obtained and to possible extensions that would enlarge the policy scope of such approaches. This concerns the existence of an optimal size, and more generally industrial composition, of regions, the links between trade and regional disparities, and the possible role of regional subsidies.

2.1 Theoretical background and estimation strategies

Maybe even more than in other fields, reduced-form empirical approaches in economic geography need to have their theoretical background clarified. Otherwise, it is very difficult to interpret the results obtained and to assess whether missing variables or endogeneity issues arise. The liter-

ature we review in this section fits into the following very simple framework. This is not really a model, but it sheds light on the relationship between productivity and the industrial characteristics of the region of interest. It assumes that a firm j operates under constant marginal cost in region r and industry s. Its output y_j depends, on the one hand, on the capital k_j and labour l_j it uses, along with its level of technology A_j and average labour efficiency s_j, and, on the other hand, on intermediate consumptions, c_j, which are made explicit here to match the data that usually concern value-added. The production function is assumed to be given by

$$y_j = \text{Min}\left(A_j k_j^{\alpha} \left(s_j l_j\right)^{1-\alpha}, \frac{c_j}{\beta}\right),$$

where $\alpha \in]0,1[$ and $\beta > 0$ are two technological parameters. Firm j's profit can be written as:

$$\pi_j = \sum_b p_{jb} y_{jb} - v_j c_j - w_j l_j - r_j k_j,$$

where y_{jb} is the quantity exported to area b and p_{jb} is the price set for region b, net of trade costs (if any) borne by the firm. The term v_j denotes the marginal cost of intermediate consumption, w_j denotes the wage and r_j denotes the cost of other inputs. Firm j's profit reduces to:

$$\pi_j = p_j y_j - w_j l_j - r_j k_j,$$

where $p_j = \sum_b p_{jb} \frac{y_{jb}}{y_j} - \beta v_j$ is the average income per unit produced by the firm, net of trade and intermediate consumption costs. With such a specification, $p_j y_j$ represents value-added, the magnitude usually provided in data sets. The first-order conditions for profit maximisation under perfect competition on both goods and inputs markets reduces to

$$w_j = (1-\alpha)\left(\alpha\right)^{\alpha/(1-\alpha)} s_j \left(\frac{p_j A_j}{r_j^{\alpha}}\right)^{1/(1-\alpha)}, \tag{1}$$

which serves as the basis for the estimation. Alternatively, one can work on total factor productivity, which is given by

$$\frac{p_j y_j}{k_j^{\alpha} l_j^{1-\alpha}} = p_j A_j s_j^{1-\alpha} \tag{2}$$

Equations (1) and (2) can be used to summarise most economic geography effects. Typically, productivity (wages or total factor productivity) is high in a region if the region benefits from technology or labour efficiency pure externalities that make A_j or s_j high. But wage and total factor productivity are also high if market access is good relatively to the degree of competition, which translates into high prices net of trade costs, p_j. Finally, firm j may also benefit from Marshallian externalities working through local input markets. If the access to inputs (or intermediate con-

sumption) is good, translating into low r_j (or v_j), then wages and total factor productivity are also high. All these effects can be read in the reverse direction. Productivity is low if pure congestion effects dominate in the region, which lowers A_j or s_j, if market access is bad - either because competition is strong or because markets are located far away (which implies a low p_j) - or if access to inputs is bad (high r_j or v_j, due to high land prices for example).

Economic geography models are much richer than the present framework. They provide micro-foundations to all market-based effects, typically working through p_j and r_j in fully-fledged models under imperfect competition and equilibrium as general as possible. To ground theoretically reduced-form approaches in economic geography, there is no need to go much further than that, however. Taking the example of the impact of density on productivity, the following wage equation is typically estimated

$$\log w_{rs} = a + \beta \log den_r + \varepsilon_{rs}, \tag{3}$$

where w_{rs} is the average wage in region r and sector s, $den_r = emp_r/sup_r$ corresponds to the total number of employees (or population) in region r (emp_r) divided by land area (sup_r) and ε_{rs} is an error term representing unexplained productivity. Typical values for elasticity β are between 0.02 and 0.09. When density is doubled, wages increase by $\left(2^{\beta} - 1\right) \times 100$, i.e., between 2 and 7% according to standard estimates. Density gaps between regions at the first and third quartile can be as high as a factor of 15, even within the same country, in which case the productivity gap is higher than 50%.

A value-added / total factor productivity equation can also be estimated, generally extended to a framework that does not assume a constant marginal cost, typically

$$\log p_{rs} y_{rs} = \alpha_s \log k_{rs} + \alpha'_s \log l_{rs} + \beta \log den_r + \varepsilon_{rs}$$

Relating these specifications to the theoretical framework implies that the assumption made is

$$\log \left[\overline{s_j \left(\frac{p_j A_j}{r_j^{\alpha}} \right)^{1/(1-\alpha)}}^{(r,s)} \right] = \beta \log den_r + \varepsilon_{rs}$$

or

$$\log \left[\overline{s_j^{1-\alpha} \left(p_j A_j \right)}^{(r,s)} \right] = \beta \log den_r + \varepsilon_{rs},$$

the left-hand-side of these equations corresponding to logarithm of the average over all firms located in region r and operating in sector s. Note first that the impact of density on productivity works through either A_j, s_j, p_j, or r_j, which means that the channel of agglomeration economies

(technology, labour efficiency, goods or inputs prices) is not identified. Second, only the total net effect of density is identified, negative effects being possibly offset by positive effects, and vice versa. Still, and as described below, the impact of density increases on productivity can be crucial from the policy point of view, making such estimations important.

Whereas local productivity is usually explained at the industrial level, s, density is not an industry-specific variable. It reflects what are called "inter-industry"' (or "urbanisation") externalities, which are the effects of the overall characteristics of the region that do not depend on s. Other proxies for inter-industry externalities that are normally included in such regressions are the overall size (land-area or total employment) 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. Recently, in relation to those more structural studies presented in section 3 below, a simple market potential variable has often been introduced to measure the role of the proximity to dense markets other than the local one. Typically, Harris' (1954) potential given by

$$\mathrm{MP}_r = \log \sum_{u \neq r} \frac{den_u}{d_{ru}}$$

where d_{ru} is the distance between r and u, is often used. This is an external market potential, own-region density not being included since its effect is evaluated separately.

The second important question this literature tackles relates to the role of local industry characteristics. This concerns externalities operating within the industry, called "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 its share in the local economy. The average size of firms, which can capture the magnitude of intra-firm economies of scale, is also sometimes considered. The last variable consists in the share of professionals in the local labour force of the industry evaluates the possible role that such workers can play as a conduit of technological spillovers.[1]

There is another reason to control, more generally, 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. Not properly controlling for the uneven spatial distribution of occupations, some regions could appear more productive even if there is no local externality, simply due to the fact that they host more skilled people. While it is interesting to understand why the distribution of skills is uneven, a question to which we return below, the literature proposes strategies to quantify the presence of local externalities separately from this composition effect of local labour markets.

Finally, this literature acknowledges the fact that local endowments, in a broad sense, do dif-

[1] See Rosenthal and Strange (2004) for a review.

fer across regions, which 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), physical geography (access to coast, river, central location, etc.) and for the quality of local institutions or the level of available technology, these last two being more relevant in the context of developing countries (including China).

Clarifying the theoretical framework as we have just done is important for an accurate interpretation of the results. It is also necessary to assess whether econometric estimates may suffer from endogeneity biases. Typically, Ordinary Least Squares (OLS) estimates assume that no explanatory variable is correlated with the residuals. In equation (3), this means that the co-variance between density and the residual is zero. There are two sources of endogeneity, missing variables and reverse causality, which we illustrate here in the case of density but which concern almost all the variables that this literature considers.

Possible missing variables are numerous. Some industries are more productive than others at the national level and simultaneously over-represented in the densest areas. This is also the case of the local endowments that increase productivity. If such variables are omitted from the regression, the effect of the density, interpreted as a presence of local externalities, is over-estimated, since it captures effects not directly reflecting agglomeration economies. Density economies can also be under-estimated if one omits from the regression local amenities (cultural goods, leisure facilities, etc.) that are also positively correlated with density. Indeed, Roback (1982) underlines that amenities attract people to cities, thereby increasing land prices, encouraging firms to substitute labour for land in the production process, thus reducing the marginal productivity of labour.

Reverse causality arises when people and/or firms choose their location according to their returns, directly linked to the local productivity. In this case, productivity shocks unobserved by the econometrician (e.g., local government choosing an efficient policy, a climate event decreasing efficiency) but observed by people and/or firms become correlated to density due to endogenous location choices, thus biasing OLS estimates.

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. If panel data are available, fixed-effect strategies can be adopted, for instance to control for aggregate effects such as the nation-wide industry-specific ones. Fixed-effects, in this case individual, can also be used to address the role of individual skills, as proposed by Combes, Duranton and Gobillon (2008). The second strategy relies on instrumentation. The aim is to find instrumental variables correlated with the possibly endogenous explanatory variables, typically density, but not with productivity shocks. One then replaces the endogenous explanatory variables in the regression by their predictor according to the instruments. As proposed for instance by Ciccone and Hall (1996),

the literature typically uses historical variables, for instance local density many decades before the years for which the model is estimated. Combes et al. (2010) also propose using instruments based on the nature of soils, which could be a determinant of very early human settlements. Because of the inertia of location patterns, both families of instruments remain correlated to current density, whereas they have little chance of being correlated with current productivity shocks. Finally, natural experiments that affect density or market-potential exogenously can be also used to identify their causal effect on productivity. For instance, Hanson (1996, 1997) uses Mexico's trade liberalisation, Redding and Strum (2008) use the division of Germany after the Second World War, although in a more structural context than those presented in section 3. Combes et al. (2011) detail all these issues related to the identification of agglomeration economies. Finally, instead of assessing the impact of local characteristics on productivity, an older literature studies the possible dynamic effects of these characteristics, i.e., their impact on local growth. This approach is described in Combes, Mayer and Thisse (2008). However, studying its policy implications, which are furthermore blurred by the weak links with theory, is beyond the scope of this paper.

2.2 Policy implications

We now turn to the possible policy implications that can be drawn from reduced-form estimations. Although some clear messages are sometimes obtained, the absence of a real objective to provide policy recommendations in this literature is striking. This is underlined by both the lack of policy-oriented interpretation of existing results towards this line and the partial inappropriateness of the strategy used.

2.2.1 How to gain more from density?

Clearly, the main outcome of this 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. It is, however, 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 (preferably skilled) 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. Typically, although one can control for the endogeneity of density, the equation relating people's location choices to the characteristics of the region has not been estimated yet. This means that the tools that could allow a policy-maker to attract more workers to larger regions have not yet been characterised in a framework consistent with the estimation of agglomeration economies. Furthermore, this should be done for each skill level separately if one wants to control the skills composition of the region's new immigrants. At one extreme, it might well be that people refuse to move to large regions in any greater numbers than at present, and the policy-maker will fail to increase the size of the region. At the other extreme, one can also reasonably imagine that people are sensitive to housing and land prices, meaning that changing land use legislation could be an effective way to influence the regional size. Again, the problem is that no quantification of the migration determinants consistent with the agglomeration economies estimation is available.

Note that the estimation of agglomeration economies also implicitly assumes that the degree of congestion of local infrastructure remains the same. For this to hold, land development would have to take place simultaneously with the increase in regional density. Typically, local transport infrastructure, schools, health services, etc. will have to be built. Obviously, such costs can be covered by taxes on productivity gains (and are netted out of the above estimates of the density impact), but the current empirical economic geography literature has very little to say about how this can be implemented and the effort it represents. In other words, 2% is only a net gain, corresponding to a $x + 2\%$ gross gain and a $x\%$ loss. Whether x is large or small, which has not yet

been assessed by the literature, can make some difference to the magnitude of the tax policy to be implemented simultaneously with the local density increase.

In the same vein, it is worth stressing that the whole discussion in this section is only about possible productivity gains from changes in the region's characteristics. Unfortunately, given the current state of the empirical literature, broader conclusions about the impact of regional size, or any other local characteristics, on the total welfare of the economy cannot be proposed. This requires the productivity gains and therefore income gains from agglomeration for the representative consumer to be weighed against the costs of agglomeration borne by the consumer, typically the impact of the region's characteristics on his cost of living. Very little is known empirically in this area; there is for instance no consensual estimate for the elasticity of housing or land prices to city size. Typically, if workers were mobile enough for their indirect utilities (income minus the cost of living) to equalise across regions, the elasticity of productivity with respect to density would also identify the impact of city size on the cost of living. There would be no welfare gain for the representative consumer from increasing city size, even if productivity increased. However, as presented in Combes et al. (2005) using a diagrammatic approach, this type of conclusion relies on a number of crucial assumptions, regarding the inter-regional and intersectoral mobility of workers, their possible differentiation in terms of skills, the magnitude of trade costs for goods, the nature of competition, the way new cities can be built and so on. This prevents us from providing a firm empirical answer to policy-makers regarding the overall welfare effects of city characteristics.

2.2.2 What are the optimal size and number of regions?

Given the log-linear form assumed for the positive impact of density on productivity, one single region of the largest possible size would maximise the overall productivity of the economy. However, this makes little sense in reality and has to be qualified from both the theoretical and empirical perspectives. One of the conclusions of the theoretical literature on systems of cities (from Henderson, 1974) states that, for each city, there exists a size that maximizes productivity (and then an optimal number of cities in the economy, generally different from one). City size increases productivity but at a decreasing marginal rate, a peak is reached and then productivity declines with city size. The peak may depend on other characteristics of the city and therefore be different for all of them. Estimating it over a cross-section of cities, or regions, can thus prove to be difficult, and further complicated by endogeneity issues like those discussed in the previous section. Beyond such problems, it is surprising that empirical specifications assume a monotonously increasing effect of density, which prevents any inference about the fact that regions are too large or too small with respect to the size that maximises productivity.

The study by Au and Henderson (2006), who focus on Chinese cities, is one of the rare ex-

ceptions that attempt at addressing these questions. China is an interesting case study, because of the country's rapid overall expansion and the presence of institutional constraints on internal migration. Both in a structural setting, in the spirit of those described below, and in reduced-form specifications, Au and Henderson (2006) find that the impact of city size on output per worker is indeed bell-shaped, implying the existence of a size for each city that maximises its output per worker. They also show that it depends on the city characteristics, in particular its composition in terms of manufacturing employment relative to services. The main conclusion reached, which is striking, is that Chinese cities are in general too small to maximise productivity. Between 50 and 60% of the cities have a size significantly below the productivity peak, while only about 2% are significantly above. Among the cities that are too small, the median productivity loss is modest, around 17%. However, it reaches 28% at the first quartile and at least 69% for 10% of the cities. According to the authors, these undersized Chinese cities are explained by restrictions on migration. Given the higher productivity they observe in cities, their intuition is that wages there should be higher, and that this difference should not be fully offset by higher goods and housing prices. Migration, if permitted by the authorities, should take place, thereby increasing city size. However, a location choices equation to determine whether such mobility will take place is not estimated, so this remains hypothetical. Moreover, even if such migration went in the right direction, one should not forget that the market outcome rarely corresponds to the social optimum in economic geography. The question of how to reach this optimum, i.e. how to induce more migration than obtained through market forces (if the market outcome is insufficient) remains.

A last issue concerns the impact on other regions of increasing the size of a region. In the absence of foreign immigration or an increase in fertility rates, the growth of one region's population can only arise at the expense of other regions. If density, and therefore productivity, increase in one region, they will both decrease elsewhere. This means that the optimal policy for a region may be different from the optimal policy at the national level, a standard policy issue when studying regional policies that is clearly present in this case.

Beyond the optimal size of each region, the question therefore arises of the optimal total number of large regions. Although largely studied from the theoretical point of view, there is, to the best of our knowledge, no empirical assessment of this issue. The difficulty lies in the fact that current specifications do not involve much of the rest of the economy. The productivity of a region depends on the characteristics of this region only and not on those of other regions (apart from the market potential variable, which is further discussed below). A general equilibrium model of regional trade would make all regions inter-dependent and should therefore be the starting point of the analysis. Typically, an experiment such as increasing the density of a region would have not only the direct, positive effect on its productivity we have just discussed but also indirect and possibly negative effects on the productivity of other regions. It might then be possible to determine

the optimal spatial allocation of employment between regions. Still, we are clearly far from having a setting allowing for all of this, and this makes it a possibly interesting future line of research.

The previous discussion focused on the role of density / overall size of the region. Similar issues arise for most other explanatory variables considered in the reduced-form specifications. Generally, the impact of own-industry size (specialisation) is positive. In France, it is around 0.02, meaning that increasing the share of an industry by 10%, increases productivity by 0.2%. Specialisation elasticities can reach 0.04 in high-tech industries or business services. First, how can a policy-maker increase the local share of a specific industry? To answer this question, one first needs to estimate the determinants of new firms' location choices. A first step in this direction is proposed by Combes et al. (2004), who estimate simultaneously (in a VAR model) the determinants of the growth of existing firms and of the creation of new firms in a region. It is typically obtained that some local characteristics that favour the latter may hinder the former. However, this remains to be further investigated. Second, one cannot increase the share of all industries at the same time within the same region. Moreover, as for density, it is probable that if specialisation in a given industry increases in one region, it will decrease in others. The optimal number of regions in the economy where a given industry is present is a question that this literature does not answer.

Another issue specific to specialisation is its possible hysteresis effect. Some researchers hypothesise that specialisation could have a positive impact on regional growth in periods where the industry expands nationally, but a negative one when it contracts. In this case, the optimal policy would depend on the period when it is applied, which can create time inconsistency. For example, increasing specialisation may be efficient in the short run but induce losses in the long run.

Lastly, one could obtain estimates of the impact of density or specialisation specific to each industry, which would enlarge the scope of policy recommendations, by making them specific to each industry. The above 2% impact of density is an average over all industries, but we can reasonably argue that it should depend on the industry, since they do not share the same nature and magnitude of agglomeration and dispersion forces. Surprisingly, such industry-specific estimates are rarely computed.

Hence, showing that local externalities on productivity exist within a framework that seriously considers selection and endogeneity biases is a first crucial step that can be achieved with current estimation techniques and data sets. However, directly deriving policy implications from current results remains difficult, mainly because empirical strategies have not been designed with this purpose in mind. We would reach similar conclusions with the literature on local employment growth, which is further confused by interpretational issues.

2.2.3 Does trade integration improve productive efficiency?

In reduced-form approaches, the productivity of a region is explained by that region's characteristics. The external market potential is the only variable that takes into account the relative location of the region within the economy. We will see below that this would be consistent with many recent economic geography models, if more sophisticated versions of the market potential variable were considered. Using Harris' market potential, as this literature does, is clearly too crude. Evaluation of the impact of trade integration on regional productivity in this context is necessarily incomplete, although it should be at the heart of regional policies. Still, a few conclusions can be drawn.

The direct positive effect of market potential that is generally obtained implies that better trade integration increases productivity. As for density, however, such a statement raises many issues. Possible selection and endogeneity biases on market potential are exactly the same as those described for density. Typically, the measure of productivity has to net out the effect of the local labour composition in terms of skills. Then, both firms and employees being attracted by high market potential regions (as they are by density), the causality may be reversed. Both issues can be addressed using the same tools (panel data and historical or geological instruments) as for density. For France, Combes et al. (2010) found the typical elasticity of productivity with respect to market potential once these biases are controlled to be very similar to the eleasticity for density, at around 0.02-0.03.

The question then arises of how to increase market potential in a region. There are only two possibilities. The first consists in increasing the size / density of nearby regions. Since market potential and density elasticities are about the same, the productivity increase for the region is the same whether density increases in that same region or in all the other regions by the same amount. Beyond the fact that the former strategy seems to be easier to implement, increasing market potential by changing regional densities is plagued by the same problems as changing own-region density. As long as the determinants of households' and firms' location choices are not characterised simultaneously with the estimate of agglomeration economies, it is difficult to determine which tools the policy-maker can use to change the density of neighbouring areas. Moreover, beyond productivity gains in a given region, the policy-maker would also have to assess the overall effect of his strategy, i.e. its impact on other regions. Typically, increasing market potential in one region can entail decreasing it in other regions, those from which the population comes. However, if one could succeed in increasing density in all regions simultaneously, typically by means of foreign immigration, all regions would benefit from both the own-density and the market potential effects.

There is a second way of increasing market potential, however. It consists in reducing trade

costs. Many components enter trade costs, typically transport costs, information costs and trade policies. We can imagine a policy-maker having some means to affect transport costs, typically by building transport infrastructure. First, the cost of building this infrastructure is not captured by the reduced-form approach, and should be deduced from possible gains. Second, some elements of transport costs, like the cost of energy, are only partly under the control of the policy-maker. He can control the corresponding taxes, but not the raw energy price. An increase in transport costs in the future can be envisioned if the cost of energy increases a lot. The policy-maker can also play on information costs, but there are very few estimates of the share of these costs in total trade costs. If one believes the few estimates of the effect of migration on trade, for instance (see Rauch, 2001, Wagner et al. 2002, and Combes et al. 2005), it could be quite large, but this remains to be confirmed. Lastly, the policy-maker can use trade policy tools, but he is clearly constrained by the political economy of such changes, both domestically and internationally. In any case, it would affect only the part of market potential that is external to the country, which is not always included in regressions.

One important issue when playing on trade costs is that this affects the location incentives of all agents, by changing the extent of competition, the prices of goods and the returns to factors. However, the impact of market potential is evaluated ceteris paribus, which is rarely the case in practice. Typically, both own- and other regions' densities change simultaneously with the fall in trade costs. Unfortunately, the magnitude of such indirect effects is not known, since, once again, no equation exists to specify how people and firms choose their location. Again, this calls for a multi-equation empirical approach dealing with both location decisions and regional productivity.

Inter-regional trade costs being rare in data sets, a further issue relates to the fact that distance alone is often used to proxy them. Combes and Lafourcade (2005) show that this is a fairly relevant strategy in cross-section. Still, it may only imperfectly tackle information costs, as shown by Combes et al. (2005), and trade policy costs. Moreover, it is useless to capture the inter-temporal variations of trade costs. While more sophisticated strategies to estimate trade costs have been considered in the structural approaches we present in section 3, this is not the case for reduced-form approaches. Therefore, the strict interpretation of an increase in market potential when regional densities are held constant is a reduction in distance, which is, strictly speaking, impossible. Moreover, the Harris' market potential used here assumes that the elasticity of trade costs with respect to distance is equal to one, while the empirical trade literature estimates a different value for the overall impact of distance on trade costs (see Anderson and van Wincoop, 2004), depending also on the particular component of trade costs (transport, information, trade policy) one wants to assess. It is therefore difficult to evaluate the amount by which one has to change distance to match a given trade costs variation when the Harris specification is used.

To conclude, it is important to note that a direct consequence of the reduced-form estimation

of the impact of market potential on productivity is that the optimal policy is to reduce trade costs as much as possible, which increases productivity everywhere. As for the conclusion that a single large region would be optimal, this shows that the model is incomplete. As noted above, there are two main concerns. The cost of reducing trade costs is ignored, and it can be large. More importantly, changing trade costs affects all the other variables in the economy and, in particular, the location choices of both households and firms, which is not taken into account. The assessment of the impact of trade costs in this framework is therefore partial. We will see in section 3 that structural approaches should prove to be more consistent in this respect.

2.2.4 Do regional subsidies affect local development?

Older strategies for assessing the role of transport infrastructure consider it as an input of the regional production function. In one of the first empirical studies of endogenous growth theory, Barro (1990) for instance proposes extra explanatory variables for government spending, among which transport can be distinguished. Such a strategy rapidly reaches a dead-end, however. The most important limit is again due to the fact that it implicitly considers that agents are immobile and do not update their location choices when infrastructure changes. More than that, since transport infrastructure clearly affects the region's trading opportunities, it modifies the region's specialisation pattern and the profitability of firms on which productivity, as defined above, depends. Moreover, since government spending in a region changes trade costs with all the other regions, it also affects productivity in those regions. This is not taken into consideration, since the region's productivity is assumed to depend only on its own infrastructure. In this respect, considering the role of infrastructure through its impact on a market potential variable would be more consistent, but this approach is ignored in this literature. Lastly, infrastructure is very often built either in expanding regions (to limit congestion of the local transport network), or, on the contrary, in depressed regions (to help them catch up). In either case, this raises reverse causality issues, since regional productivity is now a determinant of government spending. This reverse causality is rarely addressed (one reason being the difficulty of finding good instruments), Duranton and Turner (2010) being a recent exception.

In addition to the role of local government spending, there are, in many European countries, specific regional policies. Typically, subsidies are given to firms that choose to locate in certain regions, for instance to cover part of their installation costs. They can also benefit from tax exemption. Again, assessing the impact of such local policies is difficult due to the fact that they are targeted to the regions that have the lowest productivity. This is very similar to the famous example of the positive correlation between the local number of police employees and the local level of crime, due to the fact that extra police are targeted to high-crime areas. For regional policies, not

controlling for such a reverse causality could very well lead to an erroneous conclusion that subsidies have a negative effect on productivity. Among recent studies, Dallérba and Le Gallo (2008) evaluate the impact of European structural funds on the convergence process between European regions over 1989-1999, controlling for potential endogeneity. They show that although there is indeed convergence between regions, regional funds have no impact on it. Moreover, investments targeted to peripheral regions do not spill over to their neighbours. Also circumventing endogeneity, Martin et al. (2011) show that regional policy in France has little if any impact on local employment and productivity. Studying the "Regional Selective Assistance" (RSA) programme in the United Kingdom, Devereux et al. (2007) show that grants do have a small effect in attracting plants to specific areas. Interestingly, they also find that regional policy is complementary (and not a substitute) to agglomeration economies, in the sense that firms are less responsive to government subsidies in areas where there are fewer existing plants in their industry. Criscuolo et al. (2007) use an instrumental variable approach and find that the RSA has a positive effect on both employment and investment, which naïve estimators underestimate, but there is no statistically significant effect on total factor productivity. The RSA, by supporting less efficient firms, slows down reallocation to more efficient plants. Using a regression-discontinuity to identify the causal effects of Objective 1 treatment on the economic growth of EU regions, Becker et al. (2010) find positive per capita GDP growth effects, but no employment growth effects. Overall, the estimation strategy used in reduced-form approaches may not be the most relevant for testing the impact of regional policies, for the reasons developed above, and it could be improved by using the more structural approaches presented in section 3. Nevertheless, there is currently, to say the least, no strong evidence regarding the impact of regional subsidies on local development.

3 Structural approaches and the spatial efficiency-equity trade-off

We have underlined that the way trade costs enter reduced-form specifications is crude. Typically, transport infrastructure for instance is seen either as an input of the regional production function, or, at best, as a determinant of the Harris market potential. An important contribution of new economic geography models is to stress that, in general equilibrium, trade costs affect most endogenous variables, quantity, prices but also factor returns and endowments. Therefore, in empirical exercises, almost all explanatory variables 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 designed to take theory even more seriously and to use it to derive specifications directly obtained from theoretical models. This should allow the researchers to consider all channels through which economic policy variables, including trade costs, affect regional disparities.

This section follows the same organisational pattern as section 2. We first clarify the theoretical background used and the empirical specifications derived from it. While most of the empirical difficulties underlined in section 2 remain, new ones also appear. Still, we show that some policy implications can be obtained, and also that empirical strategies more specifically designed to address policy issues would be welcome.

3.1 Theoretical background and estimation strategies

Typical models used to make structural estimations are borrowed from the Dixit and Stiglitz economic geography approaches, more precisely the Krugman (1991) and Krugman and Venables (1995) settings. 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. Trade costs are of the Samuelson-Iceberg type, entering demand functions multiplicatively. 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 can deal 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, if a firm is located in or near a large region, it can access a larger market than if it is located in a small region. This is because trade costs make firms less efficient than local firms on distant markets. 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, one understands easily 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 not available in data sets.

Hanson (2005) tackles this issue with the help of Non-Linear Least Squares (NLLS). However, he needs to consider the Helpman (1998) extension of the Krugman (1991) model that considers the role of 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, which leads to a specification such as:

$$\log w_r = a + \alpha_1 \log \left[\sum_s Y_s^{\alpha_2} w_s^{\alpha_3} H_s^{\alpha_4} e^{-\delta d_{rs}} \right] + \varepsilon_r \tag{4}$$

where w_r is the wage in region r, a is a constant term, and Y_s and H_s are the total income and housing stocks in region s, respectively. The term $e^{-\delta d_{rs}}$, where d_{rs} is the distance between r and s, denotes the spatial discount factor. Hence, if we ignore the role of w_s and H_s, the explanatory variable consists in a Harris market potential[2]. The structural model shows that price effects in a fully-fledged economic geography model modify the specification and make it necessary to include other variables in the market potential function.

The approach is structural because the specification estimated is directly derived from the theoretical model and because estimated parameters (δ and the α_i for $i = 1$ to 4) 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 first econometric issue relates to the non-linearity of specification (4) with respect to the parameters to be estimated. Hanson (2005) deals with that using NLLS. Then, given that the structural approach reduces to estimating a wage equation with local characteristics as explanatory variables, the same issues regarding missing variables and reverse causality as those described in section 2 are present. They are tackled using the same tools (control variables, region and industry fixed-effects, instrumentation). The main conclusion is that such a 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 estimates are structurally consistent. They match theoretical constraints (for instance the elasticity of substitution between varieties has to be larger than one) and are close to the values generally admitted. It is also shown that the share of the spatial wage variance explained by such

[2]Except that Harris uses an inverse instead of an exponential function of distance.

a 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 (with more than two regions) model 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.

The Redding and Venables' (2004) approach is also structural but it emphasises different agglomeration and dispersion forces. In their model, the wage equation reduces to

$$\log w_r = a + \frac{1}{\alpha\sigma}\log \text{RMP}_r + \frac{\gamma}{\alpha(\sigma-1)}\log \text{SP}_r - \frac{\beta}{\alpha}\log x_r - \frac{1}{\sigma}\log(a/c) + \varepsilon_r \tag{5}$$

where RMP_r is the same real market potential variable as in Hanson (2005). The term x_r corresponds to the price of inputs other than labour, a variable which is also implicitly present in Hanson (2005). In both cases, it is assessed by control variables. Finally, SP_r is a new effect that also takes the form of a real market potential variable. It assesses the access to intermediate input markets, which depends on the location of input suppliers instead of final consumers. To tackle the estimation problems that real market potential variables create, in particular their non-linearity in unknown parameters, Redding and Venables (2004) design a strategy that uses information on trade flows. According to the model, the two market potential variables can be written as:

$$\text{RMP}_r = \sum_s \exp(\text{FM}_s)\phi_{rs} \qquad \text{and} \qquad \text{SP}_r = \sum_s \exp(\text{FX}_s)\phi_{sr} \tag{6}$$

where ϕ_{rs} is a function of trade costs τ_{rs} and FM_s and FX_s are the same regional fixed effects that enter a trade equation also derived from the model:

$$\log X_{rs} = \text{FX}_r + \log \tau_{rs} + \text{FM}_s + \nu_{rs}$$

The trade equation, which is not very data-demanding, is estimated first. It only requires the specification of trade costs as a function of distance, and also possibly as a function of other variables, some of them directly related to policies. Then, the estimates for the origin and destination fixed effects, FX_r and FM_s, are plugged back in (6) to get predictions of RMP_r and SP_r. These are finally used to estimate specification (5) with OLS.

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. Therefore, 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 one, more distant to 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 for instance richer than the US, which, from both the reality and policy points of view, is not very appealing. The second strategy uses the distances to New York, Brussels and Tokyo, which, maybe surprisingly due to their obvious correlation with current productivity shocks, pass over-identification tests. Like Hanson (2005), similar results are obtained when many control variables are introduced[3]. 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 too high correlation of the two market potential variables. Strictly speaking, it prevents one 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 areas to explain European regional GDP per capita (Breinlich, 2006) and wages (Head and Mayer, 2006), wages in Indonesian districts (Amiti and Cameron, 2007) and Chinese cities (Hering and Poncet, 2010). Lastly, like Hanson (2005), the study is performed on aggregate data, making it impossible to provide industry-specific policy implications.[4]

3.2 Policy implications

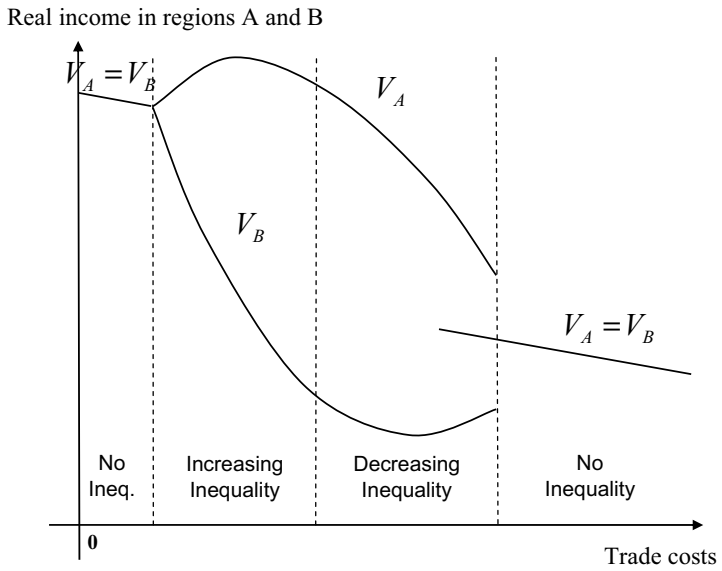
3.2.1 Does the Bell-shaped curve hold?

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. A typical pattern for disparities in regional indirect utilities of a two-region economy is represented in figure 1.

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

[4]See Combes, Mayer and Thisse for more details.

Figure 1: Trade costs and regional disparities



1

Economies that are very badly integrated do not present regional disparities. This corresponds to the right-hand side of Figure 1, where, for high trade costs, the indirect utilities V_A and V_B of two regions A and B are identical. Increasing trade integration first increases welfare homogeneously over space. In other words, efficiency is improved while perfect regional equity is maintained. However, further economic integration causes spatial differentiation, which implies a divergence between regions. In the core region (A in Figure 1) where manufacturing concentrates, a better accessibility to varieties, and possibly higher nominal wages, makes indirect utility higher than for lower integration levels, while the reverse holds in the peripheral region (B). In general, whereas efficiency keeps increasing for the average consumer, immobile workers located in the periphery lose, while immobile workers located in the core gain. This is also the case for mobile workers. This pattern corresponds to the famous spatial efficiency-equity trade-off.[5] Trade integration increases efficiency but also creates disparities. For high levels of trade integration, however, convergence occurs, with some firms relocating to the periphery, due to too high a degree of congestion in the core (competition on goods markets, high land prices and labour costs, congested transport networks, degradation of the environment, etc.). This reduces regional differentials in both access to varieties and nominal wages. The indirect utility spatial gap of immobile workers decreases, and eventually disappears, while the average consumer keeps gaining. Efficiency and equity objectives are reconciled during this phase. While based on numerous technical assumptions (regarding the nature of competition, the formulation of trade costs, the mobility of workers, etc.), this spatial pattern of regional disparities in two-region settings is considered to be a robust theoretical conclusion of economic geography models.

It is tempting to conclude that when structural estimation fails to reject economic geography models, they also fail to reject the bell-shaped pattern for regional disparities. One consequence is that trade integration is always good, for all agents if very deep integration can be achieved, at least on average otherwise. Even in the latter case, it can be shown that compensation schemes can be implemented to make everybody better off. This means that in general the gains from integration can be redistributed to the agents that it harms

These are fairly optimistic conclusions, but they are qualified by further thoughts about these issues. First, neither theoretical models nor their structural estimations are precise about which tax system should be introduced simultaneously with trade integration to achieve both efficiency and equity objectives. In particular, the above-mentioned compensation strategy is possible using

[5]See Baldwin et al. (2003, chap. 11) for one of the first systematic analysis of the efficiency-equity pattern. The literature has then proven that the conclusions reached could depend on the welfare function considered (See for instance Ottaviano and Thisse 2002; Charlot et al. 2006). Moreover, when the agglomeration mechanism entails factor accumulation or input-ouput linkages, immobile factor owners may gain and mobile factors may lose from agglomeration (see Ottaviano and Robert-Nicoud 2006; Gaigné 2006)). However, as clear from the rest of this section, the empirical literature is for the moment far from being able to empirically assess which of these sometimes conflicting conclusions are the most relevant for reality.

lump-sum transfers. However, they are difficult to implement in reality, for instance due to the presence of asymmetric or imperfect information. This requires the use of second-best incentive-compatible contracts, which can induce efficiency losses. Moreover, nothing is said about the amounts that should be transferred between agents or the levels of tax that should be levied.

Given that the perfect integration / zero trade costs situation cannot be achieved, there exists an optimal level of trade integration that does not necessarily correspond to the lowest level of trade costs. Again, it is characterised neither in models nor in their structural estimation. Typically, and on top of the nature of the welfare function used (Benthamian, Rawlsian, or other), this optimal level depends on two parameters that are very difficult to assess and can be specific to each economy: the weighting given to the different agents in the economy and the degree of inequity aversion. To summarise, a policy that consists in reducing trade costs has consequences other than average regional efficiency and must therefore be completed by other policies, typically redistributive fiscal ones. But the optimal taxation exercise in the spatial economy context faces the standard difficulties that usually characterise it and has rarely been properly and completely characterised.

The policy-maker wishing to implement spatial policies faces two further big issues. 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. 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.

3.2.2 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 and, in particular, 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[6] 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 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 monotonous 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.

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

[6] 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).

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 section 2, this clearly eliminates some channels of regional disparity, especially those self-reinforcing due to the endogenous size of regional population characteristic of the Krugman (1991) type economic geography models. Although many more general equilibrium effects are considered here by comparison to the reduced-form approaches presented in section 2, in particular the role of endogenous demand and intermediate input prices, evaluation of the impact of trade cost reduction 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 of the time horizon over which they have to be written off arises. The tools that can be used to reduce trade costs and the magnitude of the possible decrease given the current level of technology are related questions that we have already addressed in section 2. 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 inference about that, as explained in section 2 when, for instance, they 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 for instance.

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 one step 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 to change when trade costs decrease. Still, it is certainly possible to use similar settings that would allow a policy-maker to address regional disparity issues, taking into account many direct and indirect effects of his/her decisions.

3.2.3 Further regional policy issues

One advantage of structural approaches consists in their precision in the assumptions made and, therefore, in the clear interpretation that can be given to results. The drawback, when one wants to assess the impact of policies for the real world, might be that too many effects are neglected. In a way, the reduced-form estimate of the impact of density is not very sophisticated. It is only a total net effect, but one advantage is that it encompasses quite a wide range of agglomeration and dispersion effects. For instance, Hanson (2005) considers the role of crowding in the housing market, but Redding and Venables (2004) do not. Is it relevant to assess the role of a policy without considering that the spatial gradients of land price are usually large and do impact on households' and firms' location choices? Similarly, the congestion of transport networks is ignored in structural approaches, even if a direct effect of trade integration could be to increase trade flows, which would induce congestion

Another concern that should gain importance in the near future for policy-makers relates to environmental considerations. By affecting firms' and households' location choices and the volume traded between locations, regional policies have an impact on CO_2 emissions, for instance, and on environmental concerns in general. If one considers polluting emissions as a cost for future generations through their possible impact on climate and health, they should be included in the policy-maker's objective function. This would change the optimal policy. The direction of the

change is not straightforward, however. For instance, more agglomeration may imply less trade, which saves emissions, but we have seen that low enough trade costs may induce re-dispersion. Furthermore, one must take into account the movements of both goods and people, while models encompassing both inter-regional trade and local commuting are rare. The environmental impact of location choices may also depend on which sectors relocate. All of this clearly makes this line of research appealing.

Some positive effects, such as pure externalities due to knowledge spill-overs or improved matching on local labour markets, are also absent from current structural models. As long as they are used in a theoretical perspective, i.e. to better understand the various forces at work, and possibly, once calibrated, to determine the most likely equilibrium, these limits might not be a problem. But if the purpose is to provide a quantified answer for a policy-maker, the scope of agglomeration and dispersion effects considered in the model should definitely be extended. This is clearly feasible in fully calibrated models, but probably more difficult if one wants to obtain some of the parameters from a structural estimation.

Other costs that should not be ignored in a real context relate to the imperfect mobility of factors between sectors. Assuming that factors can be reallocated perfectly between sectors is standard. But this is clearly not satisfied in reality. Some costs of setting-up plants are sunk, technology and labour skills are partly specific to sectors, if not to firms, and cannot be transferred to other activities. All of this should be deducted from possible gains from the policy. There are two problems, however. The first is that mobility, be it spatial or sectoral, is in reality imperfect, whereas models assume either perfect mobility or none. This means that there is a further dimension to estimate, which is the share of the activity that is mobile. The difficulty is that this is an endogenous variable, since it depends on the degree of asymmetry between regions or sectors. The second problem relates to the fact that some of the costs of moving between regions or sectors are non-monetary. They can be purely psychological, or relate to the loss of social networks for instance. These dimensions should not be ignored by policy-makers, but are certainly pretty difficult to assess and quantify.

4 Conclusions

The conclusion of this survey could appear to be pretty pessimistic. In the end, we know little and so many dimensions are ignored from current analysis that little advice can be given to policy-makers. But the glass is also half-full.

Using both reduced-form and structural approaches, the spectrum of the policy recommendations that can be given is quite large. Reduced-form approaches allow the researchers to consider many variables that impact on regional disparities, as long as they are careful about interpretation

and endogeneity issues. Even if one is not always able to identify separately all the mechanisms at work, the overall impact of these variables can be estimated. The conclusion that agglomeration economies exist and that larger regions would make people and firms more efficient cannot be denied, even if some questions still remain to be answered, such as how to reach this goal, whether it creates extra costs not considered in the analysis, or what is the optimal number of cities in an economy. The advantages of structural approaches are the opposite. Many less issues can be simultaneously addressed, but one can be more precise about which intuitions are considered and about the underlying mechanisms and effects at work. Structural approaches make the simulation of policy scenarios possible. Generalising strategies that mix structural estimation and simulation should enlarge the scope of the policy concerns addressed by this literature. For instance, one could imagine including the role of imperfect mobility (between sectors and locations) or addressing certain environmental concerns, for instance.

Clearly, even approaches that are now standard have not been reproduced for a large set of countries and geographical levels. It would probably be useful to compare the estimates obtained, and their policy implications, between different geographical and institutional contexts. Economic geography studies on India, Asian, Latin American or African countries remain rather scarce, while the diversity of the economical context there would probably make the study of regional policies particularly interesting.

5 References

Amiti, M., and L. Cameron, 2007. Economic Geography and Wages. *Review of Economics and Statistics* 89(1), 15-29.

Anderson, J., and E. van Wincoop, 2004. Trade costs. *Journal of Economic Literature* 42, 691-751.

Au, C., and V. Henderson, 2006. Are Chinese Cities Too Small? *Review of Economic Studies* 73(3), 549-576.

Baldwin, R., R. Forslid, P. Martin, G. I. P. Ottaviano, and F. Robert-Nicoud, 2003. *Economic Geography and Public Policy*. Princeton: Princeton University Press.

Barro, R., 1990. Government spending in a simple model of endogeneous growth. *Journal of political economy* 98(5), S5-S103.

Becker, S., E. Peter, M. von Ehrlich, 2010. Going NUTS: The Effect of EU Structural Funds on Regional Performance. *Journal of Public Economics* 94(9-10), 578-590.

Brakman, S., H. Garretsen, and M. Schramm. 2004 The spatial distribution of wages and employment: estimating the Helpman-Hanson model for Germany. *Journal of Regional Science* 44(3), 437-66.

Breinlich, H., 2006. The spatial income structure in the European Union - what role for Economic Geography? *Journal of Economic Geography* 6(5), 593-617.

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

Charlot, S., and C. Gaigné,F. Robert-Nicoud, J.-F. Thisse, 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.

Combes P.-P., G. Duranton, and L. Gobillon, 2008. Spatial Wage Disparities: Sorting Matters! *Journal of Urban Economics* 63(2), 723-742.

Combes P.-P., G. Duranton, and L. Gobillon, 2011. The Identification of Agglomeration Economies. *Journal of Economic Geography* 11(2), 253-266.

Combes P.-P., G. Duranton, and H.G. Overman., 2005. Agglomeration and the Adjustment of the Spatial Economy. *Papers in Regional Science* 84(3), 311-349.

Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux, 2010. Estimating agglomeration effects with history, geology, and worker fixed-effects. In E. L. Glaeser (ed.) *Agglomeration Economics*. Chicago: University of Chicago Press.

Combes, P.-P., and M. Lafourcade, 2005. Transport costs: measures, determinants, and regional policy. Implications for France. *Journal of Economic Geography* 5(3), 319-49.

Combes P.-P., M. Lafourcade, and T. Mayer, 2005. The trade-creating effects of business and social networks: evidence from France. *Journal of International Economics* 66(1), 1-29.

Combes, P.-P., T. Magnac, and J.M. Robin, 2004. The dynamics of local employment in France. *Journal of Urban Economics* 56(2), 217-43.

Combes P.-P., T. Mayer and J.F. Thisse, 2008. *Economic Geography: The integration of Regions and Nations*. Princeton: Princeton university Press.

Criscuolo, C., R. Martin, H. Overman and J. Van Reenen, 2007. The effect of industrial policy on corporate performance: Evidence from panel data. *Mimeo* London School of Economics.

Dall´ erba, S., and J. Le Gallo, 2008. Regional convergence and the impact of European structural funds over 1989-1999: A spatial econometric analysis. *Papers in Regional Science* 87(2), 219-245.

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

Duranton, G., and M. Turner, 2010. Urban growth and transportation. *Mimeo* University of Toronto.

Forslid, R., J. Haaland, and K.-H. Midelfart-Knarvik, 2002. A U-shaped Europe? A simulation study of industrial location. *Journal of International Economics* 57(2), 273-97.

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

Gaigné, C., 2006. The 'genome' of NEG models with vertical linkages: a comment on the welfare analysis. *Journal of Economic Geography* 6(2), 141-149.

Hanson, G., 1996. Localization Economies, Vertical Organization, and Trade. *American Economic Review* 86(5), 1266-1278.

Hanson, G., 1997. Increasing Returns, Trade, and the Regional Structure of Wages. *Economic Journal* 107(440), 113-133.

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

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.

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.

Helpman, E., 1998. The size of regions. In *Topics in Public Economics: Theoretical and Applied Analysis* (ed. D. Pines, E. Sadka, and Y. Zilcha), Cambridge, MA.: Cambridge University Press.

Hering, L., and S. Poncet, 2010. Market access and individual wages: evidence from China, *Review of Economics and Statistics* 92(1), 145-159.

Henderson, V., 1974. The sizes and types of cities. *American Economic Review* 64(4), 640-656.

Krugman, P., 1991. Increasing returns and economic geography. *Journal of Political Economy* 99(3), 483-499.

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

Martin, P., T. Mayer, and F. Maynéris, 2011. Spatial concentration and firm-level productivity in France. *Journal of Urban Economics* 69(1), 182-195.

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

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

Ottaviano, G.I.P. and F. Robert-Nicoud, 2006. The 'genome' of NEG models with vertical linkages: a positive and normative synthesis. *Journal of Economic Geography* 6(2), 113–139.

Rauch, J., 2001. Business and social networks in international trade. *Journal of Economic Literature* 39(4), 1177-1203.

Redding, S., and D. Sturm, (2008). The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review* 98(5), 1766-1797.

Redding, S., and A. Venables, 2004. Economic geography and international inequality. *Journal of International Economics* 62(1), 53-82.

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

Rosenthal, S., and W. Strange, 2004. Evidence on the nature and sources of agglomeration economies. In Henderson, V., Thisse, J. (Eds.), *Handbook of Urban and Regional Economics*, vol. 4, Amsterdam: Elsevier.

Wagner, D., K. Head, and J. Ries, 2002. Immigration and the Trade of Provinces. *Scottish Journal of Political Economy* 49(5), 507-525.