2011

California Dreamin': The Feeble Case for Cluster Policies

Gilles Duranton
University of Pennsylvania

At the time of publication, author Gilles Duranton was affiliated with the University of Toronto. Currently, he is a faculty member at the Real Estate Department at the University of Pennsylvania.

California Dreamin': The Feeble Case for Cluster Policies

Abstract

This study examines the case for cluster policy. This case is theoretically ambiguous and empirically very weak.

Keywords

clusters, cluster policy, localisation economies, local economic development

Disciplines

Economics | Real Estate

Comments

At the time of publication, author Gilles Duranton was affiliated with the University of Toronto. Currently, he is a faculty member at the Real Estate Department at the University of Pennsylvania.

Review of Economic Analysis 3 (2011) 3–45 1973-3909/2011003

California Dreamin': The Feeble Case for Cluster Policies

GILLES DURANTON*
University of Toronto

This study examines the case for cluster policy. This case is theoretically ambiguous and empirically very weak.

Keywords: Clusters, cluster policy, localisation economies, local economic development

JEL Classifications: R11, R38, R58

All the leaves are brown
And the sky is gray
I've been for a walk
On a winter's day

I'd be safe and warm
If I was in L.A.
California dreamin'
On such a winter's day

John and Michelle Phillips (The Mamas and the Papas), 1965

1 Introduction

Cluster building and cluster development are now widely viewed as key pillars of local development policies. Despite countless cluster initiatives and policy reports calling for cluster strategies, the economic logic behind such policies is rarely examined beyond some tautological diagrams and the mention of prosperous clusters such as the Silicon Valley. In addition, the calls for action in the policy literature rarely quantify the alleged benefits of clustering and the costs of the policies they advocate.

This study seeks to redress this imbalance. It first examines the case for cluster policies using the basic tools of regional and urban economics. It then assesses the potential benefits from clusters using results from existing empirical studies.

*Thanks to Daniel Cohen, Philippe Martin, Henry Overman, Michael Pfüger, Frdric Robert-Nicoud, Matt Turner, and Cam Vidler for their constructive criticisms on previous drafts. This paper was funded by an institution that wished to remain anonymous after disagreeing very strongly with my conclusions and the results of other papers part of the same project.

Several findings emerge from the analysis. First, clusters do exist and their existence is associated with pervasive market failures. Inefficient outcomes are thus expected in the absence of local economic development policies. However, due to the complex nature of the market failures at work it is unclear what cluster policies should do and how they should do it. This is true even in the simplest setting. Considering richer frameworks of analysis only multiplies the ambiguities. Cluster policies can even turn ugly when implemented by less than perfectly benevolent governments.

Very large economic benefits from clusters might provide a justification for brushing aside these concerns. However, the benefits from clustering on local earnings and local productivity are small. For other outcome measures such as innovation, existing research even suggests that clustering plays a negative role.

At a more fundamental level, this skepticism towards cluster policies originates from the simple fact that the local composition of economic activity is an intermediate outcome, not a fundamental driver of prosperity. Hence, cluster strategies must overcome two major difficulties. First, clustering is not a choice variable that local policy-makers can easily manipulate. Second, this intermediate outcome is only weakly related to the final prosperity objectives that local policy-makers should be interested in.

The rest of this report is organised as follows. The next section examines the economic mechanisms behind the clustering of economic activity and uses a simple model to show why cluster policies are problematic even in the *a-priori* most favourable case. Section 3 shows that the ambiguities of cluster policies only get worse when richer settings are considered. Then, section 4 makes the case that the link between clustering and prosperity, although present in the data, is quite tenuous, making clustering a very minor determinant of local prosperity. Section 5 sketches some alternative approaches for local economic development policies. Section 6 contains some concluding thoughts.

2 A Very Simple Model of Clusters

2.1 Clusters in the Policy Literature: What is the Rationale?

The 'policy' literature on clusters is extensive.[1] A comprehensive critical analysis is beyond the scope of this study. Instead, it is best to focus on summary piece "Location, competition, and economic development: Local clusters in a global economy" by Michael Porter, the leading exponent of cluster strategies (Porter, 2000a).[2]

[1] Among many, see for instance Department of the Environment, Transport and the Regions (2000), Council on Competitiveness, Monitor Company, and Porter (2001), Department of Trade and Industry (2001a), OECD (2001), European Commission (2003), Solvell, Lindqvist, and Ketels (2003), Cortright (2006) and the many references therein. It is also possible to look at the papers of the last "Global Conference of The Competitiveness Institute" at http: //www.clusters2009.com/.

[2] Other writings from the same author on the same topic (Porter 1998b, 2000b, 2003) contain the same definition of cluster and the same framework of analysis. Altogether the four aforementioned pieces have

The oft-alleged problem with the policy literature on clusters is one of definition(s). This literature provides many definitions of what is a cluster. As noted by Martin and Sunley (2003), these definitions differ and are all rather vague. This fuzziness is certainly not helpful to the analyst trying to assess the concept. Focussing on Porter, his own definition is:

> "A cluster is a geographically proximate group of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities." (Porter, 2000a, page 16)

Following this definition a cluster, at a conceptual level, appears to be no more than the spatial agglomeration of a given 'activity' (or a "field" using Porter's words). Although it is not clear what "commonalities" and "complementarities" are, this definition points at some benefits from agglomeration. Interestingly, Porter's definition is not very different from that used by mainstream regional and urban economists since Alfred Marshall (1890). For instance, Belleflamme et al. (2000, page 161) define a cluster as a "full or partial agglomeration in one region of firms [...] that benefit from each other's presence". Within a theoretical model, this type of definition can be made precise. We show this below. Thus, *at a conceptual level*, the problem with the cluster literature may not be one of definition.[3] Instead, the problem with the cluster policy literature is one of lack of well-articulated theory: what is the 'problem' that cluster initiatives are trying to fix?

A common response in the policy literature (e.g., Solvell et al., 2003) is to argue that cluster initiatives are meant to improve local "competitiveness".[4] But a lack of competitiveness is a symptom, not the root of a problem. Put differently, improving an outcome such as competitiveness cannot be a sufficient justification for a policy. Instead, the goal of public policy should be to maximise social welfare by correcting existing inefficiencies (and inequities). To understand this difference, consider the following simple example. Faster technological progress is a desirable outcome. If firms invest efficiently, the best policy is to do nothing: Public investment will crowd out private investment and incentives for firms to invest more will lead to a waste

---

been cited nearly 1500 times in the International Citation Index as of September 2010. Despite the short time span, this count demonstrates how influential these policy recommendations have been in academic circles. As for Porter's seminal book from 1990, the citation count nears 4000 making it one of the most cited work in the last thirty years.

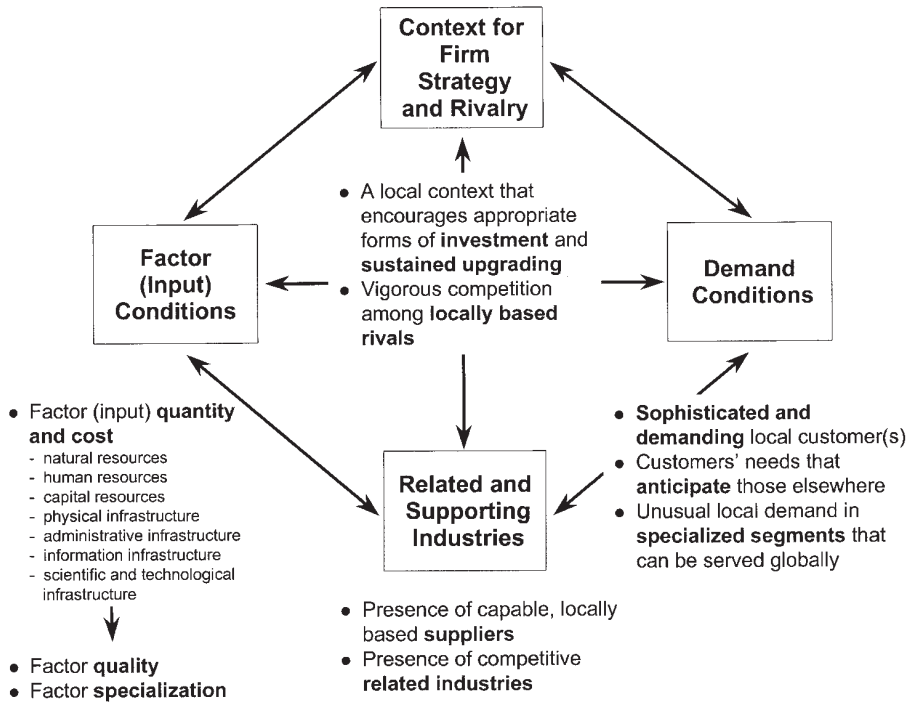
[3]Empirically, Porter's definition (just like that of Belleflamme et al., 2000) is of course more problematic since it is not obvious what an 'activity' is (in particular whether we can identify activities with 'sectors' as defined by standard industrial classifications) and what geographical scale is involved in clustering. The discussion of these issues is postponed to section 4.

[4]While economic geographers like Martin and Sunley (2003) have a problem with the definition of clusters, economists usually take issues with the use of the word 'competitiveness'. As indicated by the writings of Porter, competitiveness seems to be synonymous with both productivity and productivity growth. It could also be read as productivity relative to competition. In most of what follows, competitiveness will be taken to mean productivity in levels. Productivity growth is discussed in section 3.

of resources. If firms invest inefficiently, it is fundamental to know why. If firms do not invest enough because their innovations get quickly imitated, the appropriate policy is to increase the protection of innovations. If, instead, the absence of investment is caused by credit constrains, policy should deal with this inefficiency. More generally, any policy proposal should clearly identify some inefficiencies (or inequities) and explain how it will correct them.

To be fair, Porter in his writings is usually more concerned with articulating a rationale for his policies than most of his followers. The main theoretical tool used in Porter (2000a, page 16) and in much of his other works is the "competitive diamond" which purports to explain the "sources of locational competitive advantage" (see Figure 1). The central part of this diamond is a dynamic "local context" infused with "vigourous competition". This central element then feeds into four boxes: "Firm Rivalry", "Demand Conditions", "Supporting Industries", and "Input Conditions". These four boxes are pairwise linked through two-way arrows. The interactions between the different components of the diamond are then alleged to generate the "competitive advantage" of the cluster.

Figure 1: Porter's Competitive Diamond (taken from Porter (2000a, page 20)



According to Porter, the role of (local) governments in this framework is to foster every single part of the diamond and strengthen the arrows that join them. Because the interactions between the parts of the diamond reinforce each other, an improvement in one box gets magni-

fied through its interactions with the other components of the diamond. A number of policies are then attached to each part of the diamond. Porter (2000a) distinguishes between policies that make a cluster expand (such as "efforts to attract suppliers and service providers from other locations" and various other subsidies to increase the employment size of the cluster) from policies that make clusters more competitive (e.g., "creation of specialised education programmes", "enhancement of specialised transportation", having local authorities "acting as sophisticated buyer of the cluster's products", and a long list of other policies). Thus the final recommendation consists of many policies, all aimed at "upgrading the cluster". In turn, these interventions are alleged to result in a bigger and more competitive cluster.

The bottom line of Porter's framework is that clusters generate "competitiveness" so that strengthening and developing clusters is a worthy policy objective. In turn, this requires a complex policy mix that is specific to each cluster.

From an economic policy perspective, there are three problems with Porter's framework. First, it is unnecessarily complicated. The complexity of the web of boxes and arrows of Figure 1 is only superficial because all the elements of the model are positively feeding into each other. Consequently, the improvement of any component of the model is going to be a good thing (though it is best to coordinate improvements so that they magnify each other). There is no suggestion that a negative feedback somewhere might make things much more complicated.

Second, the complexity of Porter's framework masks the absence of crucial elements. Put differently, Porter's model of clusters is not fully specified. Then, nothing guarantees that the conclusions are warranted. For instance, nothing is said in Porter's framework about the mobility of labour. But, in absence of assumptions about labour supply, how do we know that it makes sense to attract firms from outside the cluster? If labour is sufficiently immobile, this type of policy may entail considerable deadweight losses. To take another example, there is no explicit land market in Porter's framework. But then, how do we know that the surplus generated by a cluster (if any) is not going to benefit solely passive landowners? Even worse, incompletely specified models may contain hidden contradictions. For instance, the production structure in clusters and the way firms compete are not spelled out in Porter's framework. Then, it is unclear whether the removal of barriers to entry is consistent with an increased emphasis on new product development (both advocated by Porter, 2000a, page 28). In many cases, these two recommendations are mutually inconsistent.[5] More generally, nothing guarantees that the conclusions of a model make any sense unless all the key relationships are made explicit and all the quantities are followed throughout the derivation.[6]

Third, the framework used by Porter and his followers does not really make the case for

[5]For firms to invest in new product development, they need to be able to appropriate some rents *ex-post*. Removing barriers to entry is likely to reduce these rents and thus the *ex-ante* incentive to invest.

[6]Being fully articulated does not require large numbers of variables in a model. Rather it is about making clear what the key simplifications made by the model are.

the local public policies they advocate. In a firm, if the management does nothing, chances are that nothing will happen. However, the benchmark for public policy is different. Public policy usually deals with situations in which autonomous economic agents make their own decisions, unlike passive employees who are waiting to be instructed by management. Any public policy proposal thus needs to show that it will improve on what happens in its absence. Put differently, the case needs to be made of an inefficiency (or several) that the policy will counter. This case simply is not made in Porter's analysis.[7]

2.2 Modelling Clusters

Economic theories concerned with cities, regions, or clusters have a common underlying structure.[8] This underlying structure contains three elements: A spatial structure, a production structure, and some assumptions about the mobility of goods and factors. These elements are necessary for any model to be well specified.

Spatial structure. Since clusters happen somewhere, some description of geography is obviously needed. It is often convenient to divide space into a number of areas and distinguish between the internal geography of these areas and their external geography. Internal geography is concerned with land, housing, infrastructure, and internal transport. External geography is about how areas are located relative to each other, the location of natural resources, the development of new areas, etc. Depending on the focus of the analysis, some aspects need to be spelled out in great detail while others can be modelled in a very simple fashion. For instance, models of cities often concentrate on their internal geography (land and commuting costs) and assume a very simple external geography. On the contrary, regional models usually propose a detailed modelling of the external geography of much larger areas, regions, but ignore the micro issues related to the operation of land markets. What about clusters then? Following the cluster policy literature, it is always possible to define spatial units that can be called 'clusters'.[9] Since the relative location of clusters and relative locations within clusters are not prominent concerns in that literature, we can ignore these issues and make the simplest possible assumptions about the internal and external geography of clusters. More specifically, we can think of clusters as areas that trade freely with each other and are endowed with symmetric sites.

Arguably a land market has to be explicitly introduced given the differences in land values across places and the importance of land development issues in local public policies. Hence the

[7] To be fair, many of his policy recommendations such as creating training programmes or the elimination of local barriers to entry can easily be rationalised within a consistent framework. But then the proper justifications for these policy may not bear much relationship to clusters...

[8] The material in this subsection borrows from Combes et al. (2005). Their basic framework has been adapted to the specific case of clusters as inspired by the classic work of Henderson (1974).

[9] As pointed out by Martin and Sunley (2003), this definition is somewhat tautological and may remind us of *Wonderland*'s Humpty-Dumpty whose words meant what he chose them to mean, "neither more nor less". However this issue does not matter for the time being. Its discussion is postponed to section 4.

simplest model of clusters can assume that each cluster is an area endowed with a given amount of land.[10]

These extreme simplifications may seem unappealing. It should be kept in mind however that the objective here is to present the best case in favour of cluster policies. In other words, the starting point of the analysis made here is purposefully very close to that of cluster proponents. As it will turn out, this starting point makes sense empirically. The main problem with cluster proponents is not their starting point but that they (i) do not rigourously follow through their policy argument and (ii) do not pay enough attention to the relevant magnitudes.

Production structure. It may be tempting to specify an aggregate production function that directly relates primary factors to the final output, as is customary in much of economic analysis. In the case of clusters, this standard simplification is not adequate. First, both the descriptive accounts of clusters (Porter, 1990; Saxenian, 1994) and the policy literature put a lot of emphasis on the roles of intermediate inputs, the local labour force, and non-market interactions. Second, clustering is often justified by appealing to some form of increasing returns. Understanding the inefficiencies associated with increasing returns calls for a detailed modelling of how they arise through the production process.

Three main mechanisms can be used to justify the existence of local increasing returns (Duranton and Puga, 2004). First, a larger market allows for a more efficient *sharing* of indivisible facilities (e.g., local infrastructure), risks, and the gains from variety and specialisation. For instance, a larger cluster in a given activity will make it easier to construct some dedicated facility or, for specialised input providers, to pay a fixed cost and enter the (larger) local market. Second, a larger market also allows for a better *matching* between employers and employees, buyers and suppliers, partners in joint-projects, or entrepreneurs and financiers. This can occur through both a better quality of matches between economic agents and a higher probability of finding a match. Finally, a larger market can also facilitate *learning* about new technologies, market evolutions, or new forms of organisation. More frequent direct interactions between economic agents in a cluster can thus favour the creation, diffusion and accumulation of knowledge.[11]

There is a large theoretical literature that investigates the microeconomic foundations of

[10] A more sophisticated modelling of the internal and external geography of clusters may be desirable. The very simple assumptions made here are in the spirit of the policy literature and are enough to exhibit some key inefficiencies associated with clusters in absence of corrective policies (as well as show the ambiguities of cluster initiatives).

[11] This typology differs from the traditional Marshallian 'trinity' (Marshall, 1890), which talks of spillovers, input-output linkages, and labour pooling. In fact the two typologies complement each other. Marshall's is about 'where' those effects take place (market for labour, market for intermediates, and a mostly absent market for ideas) whereas the one used here is about the type of mechanism at stake (sharing, matching, learning). Arguably, these three mechanisms (and their associated market failures) can take place in different markets. Good policies will require knowing about both the type of market failures at play and where they take place.

local increasing returns in great detail (see Duranton and Puga, 2004, for a review). Two main conclusions can be drawn from that literature. First, local efficiency (in broad sense) is expected to increase with the local size of an activity. This is consistent with the implicit assumption made by most cluster proponents that bigger and more specialised is better. Second, the sources of local increasing returns are also sources of local inefficiencies. For instance, specialist input producers in a model with input-output linkages may not be remunerated for increasing the local choice of inputs. In a matching framework, firms are not compensated for increasing the liquidity of their local labour market. With local learning spill-overs workers are not rewarded for the knowledge they diffuse. More generally, private and social marginal returns will not in general coincide in a cluster. This, of course, can justify policy intervention and may potentially be supportive of the activism of cluster proponents.

These two conclusions nonetheless come with two strong cautionary warnings regarding policy. First, the fact that many different mechanisms (sharing, matching, and learning) in different markets (labour, intermediates, ideas) can generate local increasing returns, though a very positive result (because one can assume some form of local increasing returns without having to rely on a specific mechanism to explain clustering), implies that identifying the precise sources of clustering will be difficult (Rosenthal and Strange, 2004). In turn, the appropriate corrective policies will depend on the exact mechanism(s) at play. For example, the corrective policies associated with local knowledge spill-overs are not the same as those stemming from imperfect matching on the labour market. Second, economic agents are not compensated for the positive effects they generate but they also do not have to compensate others for their positive effects. With such reciprocal external effects, it is unclear whether the wedge between private and social returns is positive or negative. For instance, a new specialist input provider may 'steal business' from competition. This implies a transfer of rents and this new entrant may thus receive more than its social marginal product. On the other hand, with knowledge diffusion social returns are likely to exceed private returns because of the lack of reward associated with knowledge diffusion. To be consistent with much of the policy literature, we can assume that social marginal returns in clusters exceed private marginal returns. One may want to take issues with this assumption and argue otherwise. Again, this is only a working assumption providing some *a-priori* rationale for pro-cluster policies.[12]

Mobility of goods and factors. Assumptions about mobility, both between and within clusters, play a crucial role. These assumptions need to cover the geographical mobility of goods, services, ideas, technologies, and primary factors. To make the strongest possible case for clusters, we can assume that some intermediate goods and services in each activity are immobile and thus

[12]Instead, should private marginal returns exceed social marginal returns, clusters would need to be taxed. Ideally, the case that social returns are below private returns should be made empirically. The *prima-facie* evidence is nonetheless supportive of the assumption made here since part of the surplus made in the cluster will capitalise in land values and thus accrue to inactive landowners.

available only to those operating in the cluster. Final goods can be taken as perfectly mobile. This is an extreme assumption which ensures that the growth potential of a cluster is not limited by the size of its market. Turning finally to primary factors of production, land is immobile (but may be used for different activities) whereas capital is arguably highly mobile. Labour and firms (and more accurately firms' knowledge) are more complicated cases. Yet again to make the best possible case for cluster policies, both are temporarily assumed to be perfectly mobile so that firms and skilled labour can be attracted to a cluster without facing strong frictions. Section 3 examines what happens when labour or knowledge are imperfectly mobile.

2.3 The Simple Mechanics of Clusters

The assumptions described above can now be brought together within a consistent model. This model can be represented diagrammatically.[13]

Productivity curve. The first relationship links the local productivity in an activity to its local employment size.[14] According to the production structure above, the more important an activity is locally, the more productive are the agents involved in it. To be more precise, the productivity of firms in a cluster increases with cluster size because of local increasing returns. The curve in Figure 2 (a) represents the productivity that firms in a particular activity face in their cluster as a function of overall employment in the activity in the cluster. For consistency with the other curves, we measure productivity in monetary units.[15] This upward-sloping productivity curve stands in sharp contrast with 'neo-classical' productivity curves. In a purely neo-classical environment, cramming more workers and more firms on a fixed amount of land results in decreasing marginal returns so clustering is a very bad idea to start with.

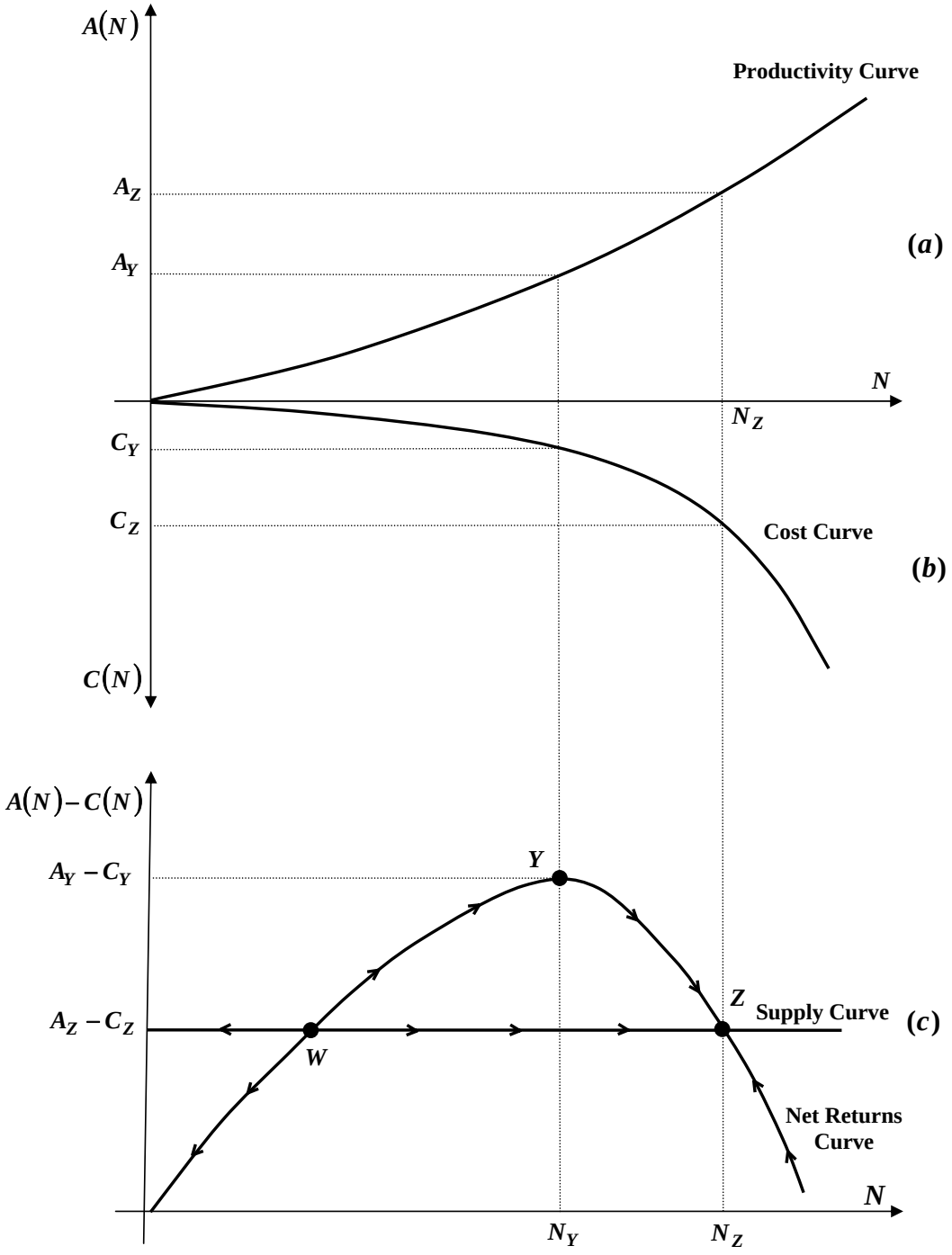
This local productivity relationship is consistent with the observed increase of most measures of local productivity with the local size of a given activity. As will be made clear below, there is considerable empirical support for such a relationship. In turn, this increase in efficiency with size can explain the 'localisation' of economic activity, i.e., the concentration of a disproportionate share of a given economic activity in a small set of areas. In Figure 2 (a), the slope of the productivity curve captures the intensity of the increasing returns in the cluster. Whether the productivity curve should be concave or convex depends on the specifics of the mechanism(s)

[13]The graphical devise used here is in the spirit of that developed by Combes et al. (2005) but focuses more on firms rather than labour to be consistent with the priorities of cluster initiatives. The 'modelling' in this report relies only on diagrams. This is to ease the exposition to the non-technical reader. A formal modelling of the ideas developed here is nonetheless needed (see Duranton and Puga, 2004, for a starting point).

[14]Employment is taken to be the key measure of cluster size for the sake of the exposition. Other measures like the number of firms could be used. The best way to measure and characterise clusters is discussed in section 4.

[15]That is, we do not distinguish between 'pure productivity' and price effects. This simplification is justified in a competitive environment and consistent with Porter's vision. See Combes et al. (2009) for a discussion of price and pure productivity effects in an imperfectly competitive setting.

Figure 2: Clusters, the Baseline Case



that underpins these increasing returns.[16] The exact shape of the productivity curve is ultimately an empirical issue. Importantly, this shape is expected to differ across activities. Local increasing returns in a textile cluster are unlikely to be driven by the exact same mix of forces as in a bio-tech cluster. These two clusters may then face two very different productivity curves. Furthermore, different clusters in the same activity may also face productivity differences that are unrelated to their size, be it only because some areas may benefit from better natural endowments than others. For instance the Northern California wine cluster, all else equal, may be more productive than its Upstate New York counterpart for obvious climatic reasons (and less obvious geological reasons).

The cost curve. The second relationship relates the local (marginal) costs of production in an activity to its local employment size. According to the spatial structure above, the amount of available land is limited in any cluster. When the number of firms and employment in the cluster grow, so does the demand for residential and commercial land. Consequently, as a cluster grows, costs are also expected to increase because of higher land prices. As an illustration, two of the largest and most successful clusters in the world, the Silicon Valley (microelectronics and software) and the City of London (finance and business services), have extraordinarily high land costs. The plain line in Figure 2 (b) represents the (marginal) costs that firms face in a given activity and cluster as a function of overall employment in this activity in the cluster. For reasons that will soon become obvious, this curve is drawn with a reversed Y-axis.

This increase in the costs of production as a function of size is modelled very simply through the crowding of the land market with a finite supply of land.[17] However, this simple relationship can be readily extended to consider a situation where more land is available further away. In this case, an increase in employment would mean a physically larger cluster with greater commuting costs, leading to higher land prices for the central locations. In turn, higher land prices have an effect on the costs of consumption goods through higher retail costs.[18]

---

[16]For instance, an ever better match between workers and firms will have some bounded (i.e., concave) benefits whereas the entry of ever more specialised input producers might lead to 'snowballing' (i.e., convex) productivity gains.

[17]An important technical issue needs to be mentioned. An increase in productivity, which raises local incomes, will have a positive effect on the demand for land and thus its price. Hence an upward shift in the productivity curve will imply an upward shift in the cost curve. This link is certainly yet another complication that cluster policies will face. However, for the sake of exposition this linkage between these two curves is ignored here. It would be possible, though cumbersome, to consider this link in a more formal model.

[18]Going against this is the fact that a larger market offers a wider variety of suppliers without having to import goods from elsewhere. Since importing goods is costly in terms of transport costs, this effects goes in the opposite direction. A higher level of activity locally may thus imply lower rather than higher costs as argued above. This effect is likely to be important and possibly dominant when one considers very large spatial units for which land issues are of second-order importance. For instance, the 'New Economic Geography' literature (which deals primarily with large spatial units) makes heavy use of this feature (Fujita et al., 1999). For smaller spatial units, however, most goods are imported so that size

Like the productivity curve, the cost relationship has also received some attention in the literature. Its precise shape is also driven by the details of the specific mechanisms that underpin it and is ultimately an empirical matter. Again, like the productivity curve, the cost curve is also expected to differ across activities since land intensity is not the same in all industries. It will also differ across clusters engaged in the same activity because of obvious differences in land availability across areas (not to mention institutional differences). Note finally that the cost curve may also reflect a range of market failures associated with the operation of the land market and commuting such as road congestion. The discussion of these complications is postponed to the next section.

The net returns curve. The difference between the productivity curve and the costs curve is represented in Figure 2 (c) by the net returns curve. On that figure, this difference is bell-shaped. This corresponds to the case where local increasing returns dominate rising production costs in small clusters, while the reverse occurs in large clusters. For this to be the case, the slope of the productivity curve must be larger than that of the cost curve below a certain threshold, while it is smaller above this threshold. At this threshold, net returns reach their peak (point Y in the figure). This peak can be interpreted as identifying a 'pseudo-optimal' cluster size, which maximises net returns in the cluster. The reason this is only a 'pseudo-optimum' (also called a constrained optimum in the economic literature) rather than a true optimum is due to the existence of inefficiencies in production. These inefficiencies imply that the productivity curve is not as high as it could be (see below for more).

The supply curve. The second curve represented in Figure 2 (c) is a supply curve for labour and firms.[19] This curve indicates for any level of net returns in the activity, the amount of labour supplied in the cluster. For simplicity, we temporarily assume that the supply of labour and firms is perfectly elastic (i.e., small wage differences imply very large migration flows). As shown below in section 3, considering an imperfectly elastic supply of labour complicates the policy recommendation even further. Since some clusters may be intrinsically more attractive because of better amenities, this flat curve may not be at the same level everywhere. In particular, better amenities will shift this curve *downwards* since workers are willing to accept lower net economic returns in clusters with better amenities.

Equilibrium. The equilibrium of the model in absence of any policy intervention can now be derived. This equilibrium corresponds to a situation where workers and firms obtain the returns they need to come and/or stay in the cluster. It is defined by the intersection between the supply

---

differences do not matter much in this respect whereas land issues become more prominent (see Combes et al., 2005, for further discussion). Hence having costs rise with local employment is an appropriate assumption in the case of clusters, which tend to be relatively small in size as argued below.

[19]For simplicity, labour and firms are thought of as being attached to each other. This could mean that there is a fixed proportion of entrepreneurs in the labour force or, alternatively, that entrepreneurs come with their workers. In reality, attracting labour (and most importantly skilled labour) and firms are two separate problems.

curve and the net returns curve. The intersection between these two curves may not be unique. In Figure 2 (c), the two curves intersect twice (at W and Z). The supply curve first cuts the net returns curve from above (at W) and then from below (at Z). As made clear by the arrows in the figure, point W is not a stable equilibrium. If the cluster is initially at point W, a small positive employment shock will raise net returns. Higher returns will then attract more firms and workers through the supply curve. This increase in population will lead to another increase in net returns and a further inflow of firms and workers. This movement only stops when the cluster reaches point Z, a stable equilibrium.

Knowing now that point Z is the equilibrium, employment in the cluster can be read at N_Z. Then, the costs of operation and productivity in the cluster can be traced upwards to Figures 2 (a) and 2 (b) at A_Z and C_Z. Before turning to policy, it remains to show that the cluster will specialise in a single activity (even in absence of policy intervention). To see this, it is useful to consider a hypothetical second activity in the cluster. This second activity has its own productivity curve and a given initial level of employment. The two activities face similar costs since firms and workers are competing for the same land. On the other hand, the two activities face, in general, different productivity levels. In turn productivity differences will imply wage differences between the two activities and workers will leave the activity with the lowest returns and move to the other. This movement ends up only when the cluster is specialised in a single activity.[20] More generally, it is inefficient to have 'disjoint' activities in the same place since they bring no benefit to each other and crowd each other's land market.

Thus far, the model shows that employment will have a natural tendency to concentrate in clusters that specialise by activity. There are some productivity benefits associated with this. These benefits are both a cause and a consequence of clustering. It is now possible to examine how the situation can be improved by policy (if at all).

2.4 What Should Cluster Policies Do?

There are two major inefficiencies associated with clusters in the framework developed here. Let us examine them in turn.

The cluster co-ordination failure. As made clear by Figure 2, the equilibrium with no policy intervention, point Z, is not efficient. The equilibrium net returns, $A_Z - C_Z$, are below those at the pseudo-optimum, point Y, which are equal to $A_Y - C_Y$. Without any corrective policy, existing clusters are *too large* with respect to their pseudo-optimum. Put slightly differently, employment in any activity concentrates into too few clusters which turn out to be too big.

This inefficiency is a co-ordination failure. All existing clusters in a given activity may be too big. However this inefficient situation can be sustained. The main reason is that no-one

[20] Should, for some unspecified reason, the two activities have the exact same returns, a small employment shock, positive or negative, in any of the two activities will again create a small asymmetry between the two activities and lead again to full specialisation for the cluster.

wants to move alone and develop a new cluster because it would mean forming a very small and thus very unproductive cluster. It would be worthwhile to move to a new cluster only if a large enough group of workers and firms decided to co-ordinate their move. Note that the creation of such new cluster would be desirable for everyone since existing clusters would become smaller and thus be able to offer higher returns. The problem is of course that, in absence of policy, there is nothing to co-ordinate this movement of workers to new clusters.[21]

Should this co-ordination failure be the only inefficiency in the framework, the policy recommendation would be clear. Existing clusters should restrict their size while empty places should develop new clusters. Although this would run contrary to the advice of most cluster proponents for whom bigger always seems to be better, this conclusion would have the virtue of being simple. Things are nonetheless more complicated because there is another source of inefficiency.

Uncompensated externalities. A second source of inefficiency stems from the production structure itself. As argued above, the microeconomic foundations of the increasing returns operating inside clusters are all associated with some inefficiency. First, the indivisibilities at the heart of sharing mechanisms generate a number of inefficiencies. Indivisibilities imply that only a limited number of players will enter. This results in imperfect competition and the (inefficient) exploitation of market power. If new entrants increase the diversity of, say, local inputs, they are unlikely to reap the full benefits of this increase in diversity. We also expect firms to make their entry decision on the basis of the profits they can make rather than the social surplus they create, which under imperfect competition is again inefficient. Second, with matching mechanisms, a different set of inefficiencies will be at play. For instance, firms neglect the positive effects of their vacancies on the job search of workers. Finally, there are also many possible inefficiencies associated with learning mechanisms: With imperfect protection of property rights, firms are likely to invest too little in knowledge generation; in absence of rewards for knowledge diffusion, too little of it will take place; firms in clusters may be reluctant to train their workers if they expect them to be poached by competition in the future, etc. These are only several of the inefficiencies that can occur when production takes place under increasing returns.[22]

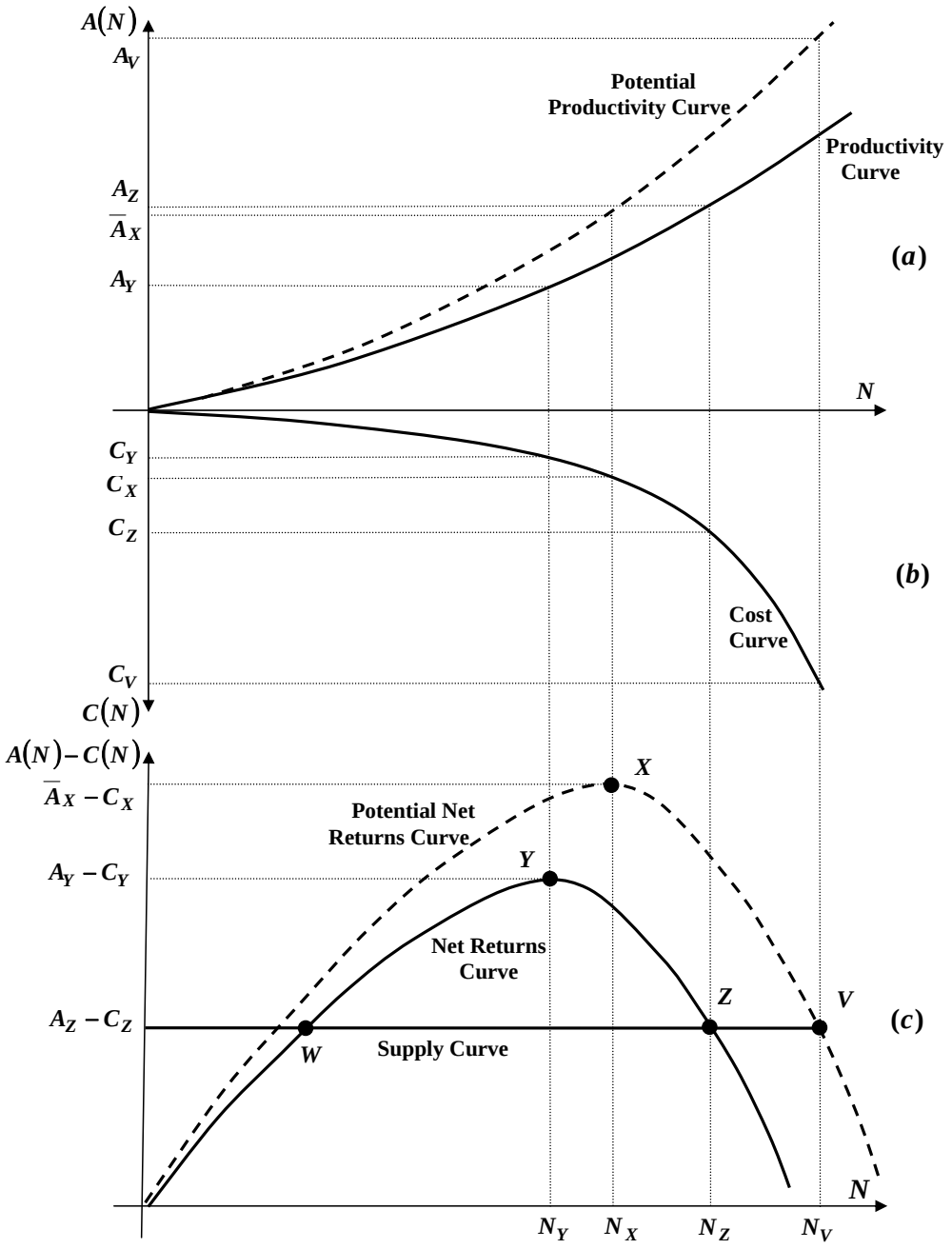
The important point to remember here is that there are possibly a very large number of inefficiencies in the production process. If these inefficiencies were suppressed, productivity would increase in the cluster at any level of employment. Using the same diagrammatic framework, this suggests that there is a potential productivity curve that the cluster could attain if all the inefficiencies in production were corrected. This potential productivity curve is of course above the productivity curve used previously. It is represented by the dashed line in Figure 3 (a).

Subtracting the costs curve (in Figure 3 b) from the potential productivity curve yields the

[21] An alternative to government policy would be to rely on the existence of profit maximising 'large agents' as in Henderson (1974). This option is not available in most countries.

[22] Duranton and Puga (2004) examine a much longer, but still very incomplete, list.

Figure 3: Cluster Policy



potential net returns curve, i.e., the net returns that could be attained in absence of inefficiency in production. It is represented by the dashed line in Figure 3 (c). On this figure, potential net returns are maximised at $\overline{A}_X - C_X$ for a level of employment N_X (point X).

Before discussing policies in greater detail, it is useful to summarise the model. Clusters occur because of local increasing returns. Despite increasing returns, clusters cannot grow infinitely large because of local crowding. Under some reasonable restrictions regarding the shape of the productivity and cost curves, net returns in the cluster are bell-shaped. There are two major inefficiencies. First, local increasing returns will rely on some uncompensated external effects and thus be associated with some inefficiency in production. Second, because of increasing returns, economic agents will also face a co-ordination failure.

Thus two sets of policies are needed for a cluster to maximise its net returns. First, the coordination failure needs to be solved so that employment in the cluster is equal to N_X. For existing clusters, this implies restricting their employment size in absence of intervention from N_Z (the level of employment associated with the equilibrium Z in absence of intervention) to N_X. Alternatively, for places that did not have any activity at the no-intervention equilibrium, this means creating a cluster of size N_X from scratch.[23] However, this alone is not sufficient since this would bring the net returns to the level corresponding to N_X on the plain curve rather than the optimum at $\overline{A}_X - C_X$. To reach this point, the policy must also cure all the inefficiencies in production so that economic agents face the dashed productivity curve in Figure 3 (a). Solving only the inefficiencies in production without dealing with the co-ordination failure would not be satisfactory because the increase in productivity would only attract more firms and workers. The cluster would then reach a new equilibrium with employment level N_V and the net returns would be the same as without intervention ($A_Z - C_Z = A_V - C_V$).

All this suggests that any cluster policy should use two types of instruments. The first should aim at fixing the inefficiencies in production while the second should restrict cluster size so that it reaches its optimal size.[24] A recommendation such as "Make production efficient and reach point X" is, of course, easier said than done.

The first major difficulty is the precise identification of the inefficiencies in production so that the potential productivity curve is reached. To repeat, clustering is driven by a variety of mechanisms whose relative importance is extremely difficult to identify empirically. These mechanisms then all call for different corrective policies. For instance, corrective policies aimed at dealing with labour market matching problems will have nothing to do with those aimed at fostering knowledge diffusion, etc. Put differently, we need some corrective policies for

[23]See next section for a discussion of the practical complications associated with implementing this option.

[24]The idea that clusters may be too big is not so fanciful. Going back to two examples mentioned above, the City of London and the Silicon Valley, welfare in these clusters would arguably increase if they saw a reduction in their crowding.

inefficiencies that we know close to nothing about.[25] The second major difficulty is to locate point X. Since the level employment associated with maximised returns depends on a curve that is unknown (the *potential* productivity curve) and a curve that is possibly very difficult to estimate (the cost curve), it very hard to know for which level of employment the difference between the two curves will be maximised.[26]

In conclusion, the policy recommendations derived here differ quite significantly from those of cluster proponents despite very similar starting points. There are two main differences. First, the vast array of policies that are often recommended for cluster upgrading by cluster proponents seem only vaguely related to the market failures associated with clustering. Some of the recommendations of Porter (2000a) such as specialised training, university research efforts, or specialised transportation infrastructure are very expensive items. This may be a lot of money spent for unlikely solutions to ill-identified market failures.[27] Second, recall that Porter's (2000a) objective is to increase both size and productivity (or, equivalently, increase productivity both directly and indirectly through a larger size). It is as if the model he had in mind was only the top part of Figure 3 and the objective was to move rightwards along the productivity curve as well as somehow shift it upwards. As made clear by parts (b) and (c) of the figure, this is not an adequate policy.

3 Clusters in a More Realistic World

Despite using a very simple framework, the previous section has shown that cluster initiatives as often advocated have a very ambiguous rationale. When rigorously followed through, the policy recommendations derived from cluster theory seem far less appealing than cluster proponents would like us to believe. As shown in this section, the problems only worsen when one tries to make the theoretical framework more realistic.

3.1 Imperfect Mobility and Knowledge Replicability

In the model described above, an upward-sloping productivity and a downward-sloping cost curve lead to a bell-shaped curve for net returns. These are probably reasonable first-order approximations of the empirical reality of clusters.[28] The assumption of a flat supply curve

[25]Furthermore, these inefficiencies are likely to occur in most places. The fact that they are not dealt with by central governments is possibly an indication of how difficult they are to solve and not a good reason for local authorities to attempt to cure them.

[26]There is a further complication caused by the fact that cluster creation in a sector increases its output and thus changes relative prices in the economy. This means that cluster policies shift the productivity curve. This makes it particularly hard for any given cluster to know where optimal size will be.

[27]This statement does not mean that training or higher education should not be of concern to policy makers. These interventions are criticised here as tools to foster clusters.

[28]Empirical results are discussed in greater detail in the next section. Direct evidence about net returns to size being bell-shaped is provided by Au and Henderson (2006) for Chinese cities.

for labour and firms is potentially more problematic. Labour may be fairly mobile in North America but it is far less so in Europe and elsewhere. Among others, Obstfeld and Peri (1998) estimate that the elasticity of regional migration flows to wage differences is relatively high in the US but extremely low in Europe.

Turning to firms, the evidence is more complex and also more patchy. Direct evidence from Duranton and Puga (2001) who report establishment relocation data for France shows that about 1.5% of the stock of establishments relocate every year. Although nearly three quarters of relocations are towards clusters, the proportion of relocating establishments is arguably low. Pellenbarg (2005) reports slightly higher numbers for the Netherlands but most of the movements he reports are very short distances. Further (indirect) evidence about the low mobility of firms is provided when estimating the effects of subsidies to locate in disadvantaged areas. According to Devereux et al. (2007), a grant of £100,000 only increases the probability of relocation to an area by 3% (for instance from 1.00% to 1.03%). Although systematic direct evidence is still missing, the literature is suggestive of low rates of geographic mobility of existing firms.

There is, however, evidence of considerable mobility of economic activity through entries and exits and the growth and decline of existing firms.[29] Hence employment in many activities may be fairly mobile by means of entries, exits and firm growth. Nonetheless, we do not know to what extent this mobility is related to differences in net returns.

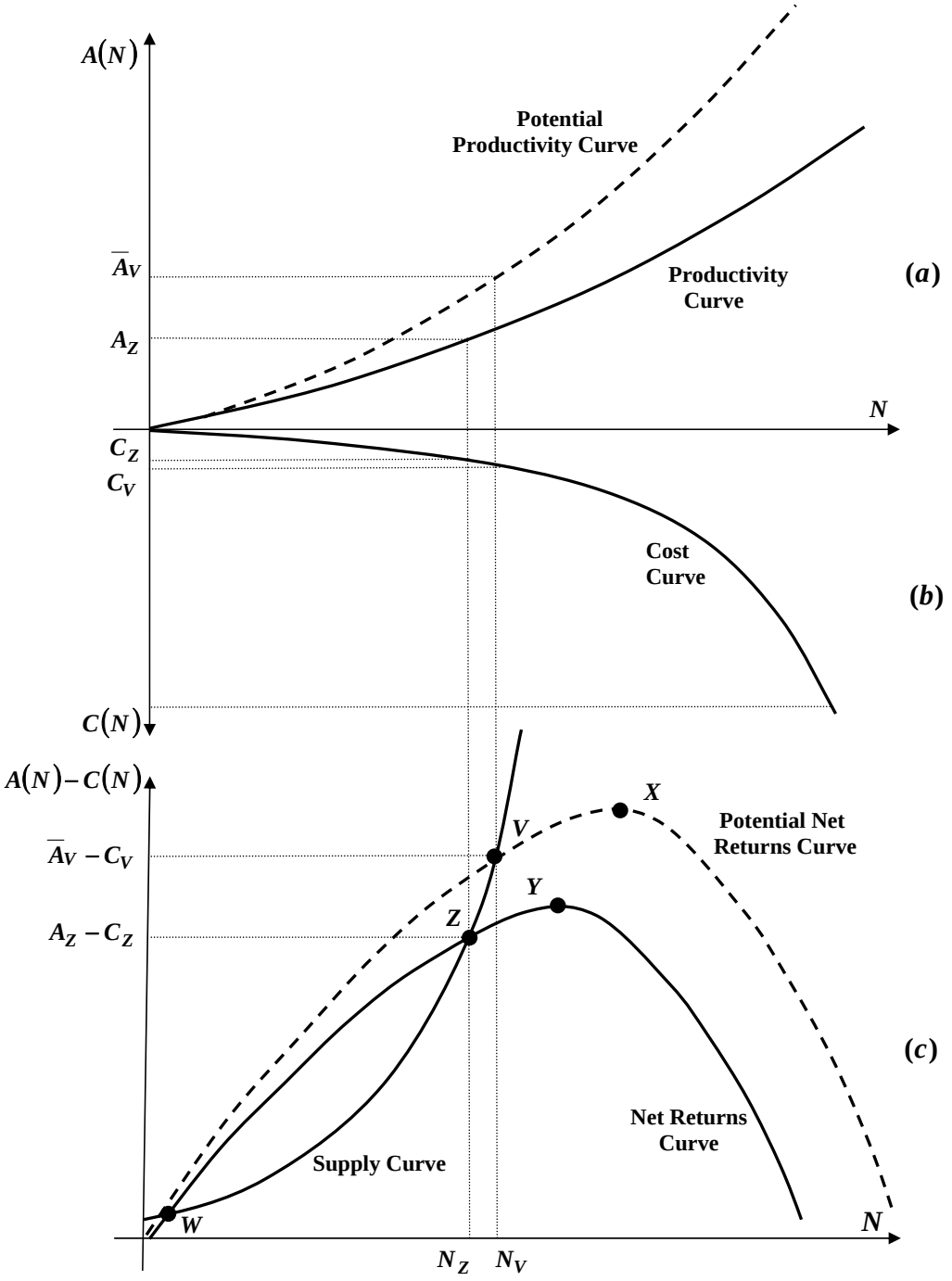
Put together, these considerations imply that the supply curve is highly unlikely to be flat as in Figures 2 and 3. Instead of perfect mobility, it is preferable to assume imperfect mobility of factors.[30] In turn, imperfect mobility is associated with an upward-sloping supply curve, which could possibly very steep in some places. Figure 4 (c) represents such a case. In this figure, the supply curve intersects with the potential net returns curve from below before its peak.

Under these conditions the policy recommendations regarding clusters are fundamentally altered in several ways. First, there is no co-ordination failure leading to clusters being too big. Dealing with the co-ordination failure and the other inefficiencies of the model already lead to a complex policy recommendation in the baseline case. Acknowledging now that the co-ordination failure that makes clusters too big may or may not be present adds another layer of uncertainty. The second fundamental difference with the baseline case regards the desirability of attracting firms and labour from outside the cluster to reach the maximum net returns at point X. When labour and firms are highly mobile it may make sense to propose some subsidies to attempt to attract more workers and firms. For the low mobility case represented in Figure 4 (c), point V is the best the cluster can hope to achieve after having corrected the market failures

[29] See Caves (1998) for a general survey on firm turnover and churning; Faberman (2005) for the spatial patterns associated with this turnover; and Duranton (2007) for summary evidence about sectoral changes in cities.

[30] Alternatively one could think of supply curves as being upward-sloping in the short-run and flat in the long run. The low levels of firm mobility documented above suggest that the 'long run' is well beyond the time horizon of most existing policies.

Figure 4: Clusters with Imperfect Mobility



in production. Attempting to reach point X makes no sense since the cost of the subsidy would outweigh the benefits from higher net returns.[31]

The third complication is that the supply curve itself may be subject to some market failures. Put differently, there may be some inefficiency that make the supply curve too steep (or too flat). For instance, imperfect information about wages and cost of living can make risk-averse workers reluctant to move to a cluster. Although the empirical evidence concerning market failures associated with internal migration is still very sketchy, there is strong evidence of regulatory barriers to mobility (Greenwood, 1997). The myriad of small (and sometimes big) barriers to internal mobility that have been documented include public housing when allocated by a queuing system, the workings of unemployment agencies, the high taxes when buying or selling a house in some countries, etc.

To summarise, considering that firms and workers are imperfectly mobile introduces three complications: (i) the co-ordination failure that makes clusters too big may be absent, (ii) offering subsidies to attract firms and workers may be pointless and (iii) the mobility of firms and workers may be subject to market failures. In addition to this, local governments should be wary of dealing with market failures associated with the mobility of firms and workers since what may lower the cost of attracting workers to a cluster may also lower their cost of leaving that cluster. On the other hand, central governments may face large efficiency gains from improved internal mobility. Despite 2006 being the "European Year of Worker Mobility" for the European Union, European governments are very reluctant to take any step to reduce existing barriers to mobility. A simple reason behind this lack of action is the fact that increasing mobility is likely to create both winning and losing clusters. If a steep supply curve like that in Figure 4 is made much flatter (as in Figure 2), this raises efficient cluster size from N_V to N_X. If total employment in the activity remains unchanged at the aggregate level, having bigger clusters also means having fewer of them.

This brief discussion does not aim to be a complete analysis of mobility issues. It is only here to show that clustering also bears some important relations with broader policy issues, which are possibly best dealt with at the level of central governments. These issues are ignored by cluster proponents who usually mention attracting leading firms and highly skilled workers as a routine exercise.

3.2 Congestion and Other Frictions

After observing the complications induced by a more realistic supply curve, we can now turn to those associated with a more realistic approach to the cost curve. In the baseline case, costs are simply assumed to increase with cluster employment following the crowding of the local land market. The implicit assumption made in the baseline case is that private marginal costs (those

[31]The deadweight loss of getting from point V to point X is the area below the supply curve and above the potential net returns curve. In the present case, it is likely to be very large.

paid by producers) were equal to social marginal costs (the costs to the economy). With a perfectly functioning land market and in absence of congestion, this equality between private and social marginal costs holds naturally. Empirically, we expect neither of these two assumptions to hold. Land markets are subject to significant frictions and are strongly regulated through planning and zoning regulations. Congestion in local transportation is also pervasive when clusters start growing.

A first implication of congestion and frictions on the land market is that the cost curve in the absence of policy is distorted. With proper corrective policies, it should be possible to reduce costs for any level of employment in the cluster. For instance, a congestion tax could reduce the level of congestion in the cluster and increase total surplus. In the diagrams used so far (e.g., Figure 4), this suggests that we should also consider a potential cost curve that stands above the cost curve in absence of policy intervention (keeping in mind that the vertical axis is reversed in the middle part of the figure). Potentially lower costs would then mean an even higher potential net returns curve in the lower part of the figure. Depending on the shape of the potential cost curve relative to the cost curve in absence of policy, optimal cluster size may increase or decline. Hence, considering some obvious complications associated with the imperfections of the land market and commuting generates yet further complications for cluster policies.

Second, there is another subtlety that cluster proponents do not appear to acknowledge: Cluster policies may conflict with other local public policies.[32] Porter's motto involves "removing obstacles, relaxing constraints, and eliminating inefficiencies" (Porter, 2000a, page 26). In turn, this leads to a vast array of policy measures to improve "competitiveness" (see for instance Porter, 2000a, page 30–32). Assuming that the absence of any mention of the land market in the writings of Porter is an oversight, "removing obstacles" and "relaxing constraints" should also apply to local zoning and planning regulations.[33] Scrapping planning and zoning regulations may be harmful. Although the effects of zoning and planning regulations are intensely debated in the literature (see Fischel, 2000, for an introduction), a case can be made that they can improve the quality of life for local residents. For instance, although the land use regulations in the Bay area around San Francisco appear suboptimal to many observers, the quality of life around San Francisco would in all likelihood decline if all "obstacles" and "constraints" on land use were to be removed. Put differently, efforts to strengthen clusters may clash with other other existing policies and cluster proponents do not offer any guidance about how to resolve these conflicts.

[32]Distributional issues are also ignored by cluster proponents. A successful cluster policy is unlikely to create only winners. Residents on fixed incomes will lose from any increase in local land prices associated with cluster growth.

[33]This absence is conspicuous in the writings examined here. There is a similar absence in Solvell et al. (2003) and most of the cluster policy literature. The report by the UK Department of the Environment, Transport and the Regions (2000) is an exception though it approaches land issues very differently.

3.3 Cluster Policy in a Dynamic Setting

The analysis so far has been mostly static in nature. This contrasts with the emphasis of Porter and many of his followers on the dynamic dimensions of "competitiveness" such as innovation and productivity growth. However, it can first be argued that much of this prominence given to the rate of growth is unwarranted. To understand why, it can be noted that increasing the annual rate of growth of local productivity by, say, one fifth would constitute a very successful productivity growth policy.[34] With productivity growing at around 2% per year (and often less), this enormously successful policy would only mean an extra 0.4% per year. For the cumulative effects of such policy to become sizeable and generate a gain in levels of, say, 10%, one would need to wait for nearly 25 years. In contrast, reducing urban congestion or improving local infrastructure may have an equally large effect within months rather than years. So, despite the insistence of cluster proponents on dynamic aspects, static effects as discussed above do matter.[35]

The second response is that considering explicitly dynamic issues introduces yet further complications to cluster policies. The life-cycle of clusters is the fundamental issue here. It begs two key questions. First, can clusters be created from scratch? Second, what should a cluster policy be like if clusters may come and go?

To answer the first question, the policy literature is divided. Porter clearly states that "there should be some seeds of a cluster that have passed the market test before cluster developments efforts are justified" (Porter, 2000a, page 26). Many cluster initiatives are not as cautious and often propose to develop new clusters from scratch. Software, microelectronics, and biotechnology are usually the industry of choice.

Empirically, Porter's caution seems to be warranted. Even though the knowledge-base of a cluster might be expanded, it cannot be created out of nothing. This difficulty of replicating clusters can be explained by the fact that cutting-edge knowledge at the technological frontier in most activities is mainly tacit and spread across a large number of workers. This immobility of cutting-edge knowledge (at least in the short run) has been evidenced in a number of ways. For instance, the literature on patent citations finds strong localisation effects whereby new patents tend to rely disproportionately on previous innovations that were developed nearby (see section 4 for more on this). In a very different vein, van der Linde (2003) finds evidence of only one case of successful creation of a cluster *ex-nihilo* by policy (Hinshu in Taiwan) in a survey of more than 700 clusters. One of the 'fathers' of the Silicon Valley, Frederick Terman, who as

[34]Despite productivity growth being the major focus of UK economic policies since 1997, the evidence of faster productivity growth in this country is yet to materialise. See for instance the chapters in Department of Trade and Industry (2001b).

[35]Cluster proponents may respond that what they call "dynamic effects" would encompass any improvement in productivity rather than only improvements in the rate of productivity growth. In such a case, however, the discussion of the previous section can no longer be dismissed for being "static". Furthermore, giving such a broad meaning to the word "dynamic" would make it useless as a qualifier.

dean of engineering and provost of Stanford University oversaw the creation and development of the Silicon Valley, was unable to replicate this again in the highly favourable environment of New Jersey some years later when poached by Bell Labs (Leslie and Kargon, 1996).

However, the fact that today's cutting-edge knowledge is not mobile does not imply that clusters will remain what they are and where they are forever. The reason is that the current frontier of knowledge will eventually be pushed forward and the creation of a new generation of cutting-edge knowledge may happen elsewhere. When George Eastman revolutionised the photographic industry in Rochester (NY), the latter city took over New York as the main cluster for the industry. After the emulsion coating technology was in turn made obsolete by the digital revolution in photography, Rochester lost much of its prominence in this industry. The systematic evidence in Arzaghi and Davis (2005) and Duranton (2007) shows that this example is not an isolated case but representative of a pattern. Activities move around, sometimes quite a lot. Worse, the more high-tech and clustered industries (often cherished by cluster proponents) seem to be even more mobile Arzaghi and Davis (2005). Hence, increasing clustering means increasing the risk exposure of the local economy. Besides, trying to glue activities somewhere may be futile (as was Eastman-Kodak's attempt to resist the digital revolution in photography). This resistance may even be socially harmful since it can slow down the development of better alternatives elsewhere.[36] A good case can be made that great cities are great not because they have managed to keep their leadership in one activity since the dawn of time but because they have managed to periodically reinvent themselves after loosing an important part of their economic fabric. Boston is a case in point (Glaeser, 2005).

In summary, taking a more dynamic perspective complicates the problem of cluster policies even further by underscoring uncertainties about how long an activity can stay in a given cluster.

3.4 When Cluster Policies Can Turn Ugly: Political Economy

The argument so far is that even if the public authority that oversees the cluster is highly competent and attempts to maximise local welfare, an optimal cluster policy looks like something extraordinarily difficult to achieve. Now, two further questions need to be addressed. Are those in charge of cluster policies really able to pursue them competently? And if so, are they trying to maximise social welfare? To answer the first question, it should be noted that in most countries cluster policies are conducted in a highly decentralised manner by sub-national governments or by some mix of sub-national government and semi-autonomous bodies representing the actors of the cluster. In most cases sub-national levels of governments do not have the capabilities to design complex economic development policies. They naturally call upon consultants to help

[36] At the early phases of their development, alternative technologies and designs are often not obviously superior to existing products. For instance mechanical calculators were produced for many years after the development of the first electronic calculators. Trying to strengthen the mechanical calculator cluster around the London in the 1970s would have been a mistake for the London region and might have harmed the development of a new industry that tremendously reduced the cost of basic computations.

them. The shortfalls of the conceptual frameworks used by these consultants are not encouraging for the possibility of successful cluster policies. Turning to the second question, there is a very large political economy literature that shows how and why governments cannot be expected to be fully benevolent (see the recent synthesis of Besley, 2006).

The fact that the institutions in charge of cluster policies may not be entirely competent nor fully benevolent may seem like a truism since these two caveats apply to most, if not all, policies. They are, however, especially important in the case of cluster policies. A lack of competence can be expected to matter more for complex policies such as cluster initiatives. A lack of benevolence is also especially damaging here. This is because when it comes to clusters, there is a double asymmetry of information. The producers in a cluster will know more about their activity, their level of performance, and the local prospects than the authority in charge of cluster policies. In turn, the local policy-makers will know much more than their voters. This double asymmetry of information can be particularly detrimental. In the worst case scenario, a group of industries can collude with their local authority and use a cluster initiative as a way to extract resources for themselves. A less extreme version suggests that producers may be able to mislead policy-makers and that this can go un-noticed by the voters. In some cases it may also be hard for elected officials to resist strong lobbying by an important group of local industries and the promise of great future wealth. On the other hand, it will be extremely difficult for voters to evaluate the benefits of a cluster policy. For instance, the lay-voter will find it much more difficult to assess a dysfunctional cluster initiative than substandard garbage collection.

The asymmetries of information associated with clusters will also affect behaviours in a variety of other ways. Politically, it is 'better' to focus on more 'visible' projects. Hence, between a number of small and potentially beneficial local initiatives and something more shiny like trying to attract bio-tech firms, the political choice is often biased in favour of the latter, however low its probability of success. Producers in badly performing clusters may also find it worthwhile to invest in convincing local authorities that they need a cluster initiative rather than in the development of their own productive capabilities (Baldwin and RobertNicoud, 2007). The list of political economy problems could be much longer. In conclusion, cluster policies that already look fraught with difficulties in a world of benevolent governments look extremely unappealing when political agency is explicitly taken into account.

4 Does Clustering Make Clusters Prosperous?

The last two sections showed that cluster policies may have some payoff but their design and implementation are much more complicated than cluster proponents would like us to believe. Furthermore cluster policies can easily be captured by special interests. These concerns may be brushed aside on the ground that even an imperfect cluster policy could still provide large economic benefits. Thus, it is time to look at the facts regarding the effects of clustering.

4.1 How to Measure Clusters

To estimate the benefits of clustering, two key measurement issues need to be addressed. They regard the industrial and spatial scope of clusters.

The theoretical sections above talk of 'activities' without trying to be more precise. Porter is equally vague when he talks of 'fields'. Not only is he vague (like all other cluster proponents) but he insists on the idea that "cluster boundaries rarely conform to standard industrial classifications systems, which fail to capture many important linkages across industries" (Porter, 2000a, page 18). This point is underscored over and over again in the cluster policy literature. A roughly similar point is made by Porter and his followers about the spatial scope of clusters, which, they argue, do not conform to the boundaries of customary spatial units such as states, counties, or metropolitan areas.

These two points made by cluster proponents are undoubtedly true. We do not expect economic activities to tie in very neatly to predetermined industrial and spatial categories. Nonetheless, these two points are mostly unimportant when it comes to assessing the effects of clusters. The first reason is that most of the examples that are discussed in the cluster literature actually correspond to narrowly defined industries in regions or metropolitan areas.[37] These narrowly defined industries seem to be well captured by the standard industrial classifications while their cluster locations are roughly consistent with existing regional and metropolitan boundaries. For instance, the list of the 41 most important traded clusters in the US listed in Porter (2003, page 565) reads like the North American Industry Classification System: Business services, Financial services, Hospitality and tourism, Education and knowledge creation, and Distribution services, to name only the first five on Porter's list.

The second reason is that assessing the benefits of clustering does not require activities in clusters to coincide exactly with standard industrial categories. It only requires clustered activities to be well proxied by standard industrial categories. To see this, note first that the standard method to assess the benefits of clustering involves regressing some local outcome by sector, such as output per worker or wages on some measures of local employment in the sector (more on that soon). Provided the measurement error on the latter is uncorrelated with other explanatory variables, a correct estimate will be obtained.[38]

---

[37] Some clusters may span much bigger areas. For instance, cluster proponents sometimes want to put the cluster tag on very large areas like the US manufacturing belt or some component of it. Sometimes entire countries are even labelled as 'clusters' (Porter, 2000a). This may be a stretch. The main issue is that another theoretical framework putting a lot of emphasis on trade linkages between firms across large distances is needed to think about such "clusters". The policy implications of this alternative framework then differ from what we described above. For more on this issue and, more generally, on spatial policies at a larger spatial scale, see Martin (1999), Puga (2002), Ottaviano (2003), and Combes et al. (2005).

[38] Although a correlation between cluster size and mis-measurement is conceivable (e.g., standard industrial categories capture cluster employment better in small clusters than in big ones), it is hard to imagine large biases here.

It is also possible to use input-output matrices to control for the sectoral linkages underscored by cluster proponents. This is what Dumais et al. (1997) propose in their work. For US counties and states, they measure the size of the local market that a sector faces by summing the size of all the sectors locally and weighting them by their share of inputs bought from that sector. For instance, leather producers face a large market locally when there are many footwear, apparel, and furniture firms around them. A mirror concept can be defined on the input side. This approach certainly captures Porter's definition of a "geographically proximate group of interconnected companies" very well. Arguably, the extensive nature of input-output matrices allows research to capture these linkages more comprehensively than the often arbitrary definitions used by cluster proponents.

Dumais et al. (1997) actually go beyond the notion input-output linkages since they also construct detailed measures for labour market pooling and technological spillovers at the level of each sector in each area. More specifically, for labour market pooling they use a matrix of occupations by sector. A given sector faces a thick local labour market when it and other sectors that employ a large fraction of their workers in similar occupations are locally large. For technological spillovers, the matrix is constructed from patent citations (using the sector of origin of the citations of the patents by sector). Many interesting results emerge from this work. The main ones regard the importance of local labour markets versus technological spillovers and input-output linkages to explain clustering effects and the spatial scale at which they matter. Interestingly, this analysis also shows that the magnitudes of clustering effects are not greatly affected when using these sophisticated measures of agglomeration instead of much cruder ones such as local sectoral employment.

What about the spatial scope of clusters? The fact that clusters do not coincide exactly with existing spatial units is another source of measurement error. The early literature on clustering effect was aware of this problem and a simple solution is to run the same regressions at different spatial scales. Similar magnitudes were typically found for relatively small units such as US counties and relatively big units such as US states. Recent comprehensive evidence by Rosenthal and Strange (2001) suggests some attenuation of clustering when one considers gradually larger units (by moving from zipcodes to US states). Rosenthal and Strange (2003) confirm this attenuation result by estimating agglomeration effects at a very fine spatial scale and allowing for spatial lags. Taken together, these results suggest that it is appropriate to use fairly small units like counties or slightly larger ones like metropolitan areas and that doing so is unlikely to result in big measurement errors.

These results are confirmed by two radically different approaches. First, in their analysis of clustering in UK manufacturing industries, Duranton and Overman (2005) use continuous space so that their results are independent of spatial units. They find that about 50% of industries indeed show some tendency to cluster. Interestingly, clustering usually takes place at short and intermediate distances (i.e., between 0 and 50 kilometres). This suggests that metropolitan areas

are the most adequate spatial level at which to examine clustering in the UK. Furthermore, the clustering of industries is most apparent at the four-digit level suggesting that clustering is best explored using finely-defined industries. In further research, Duranton and Overman (2008) also look at the co-localisation of industries related through input-output linkages. Consistent with the assertions of cluster proponents, they find that vertically-related industries do have a tendency to co-locate. However, this co-location takes place at a larger spatial scale (up to 120 km) suggesting that the regional level might be most appropriate in this case.[39]

Second, a detailed analysis of spatial measurement errors is performed by Briant et al. (2010). They estimate the benefits of agglomeration for France using a number of existing spatial units some of which are purely administrative (e.g., dpartements) while others were defined by economic principles (e.g., employment areas). They also perform the same exercise for entirely arbitrary units (i.e., squares overlayed on a map). Their main finding is that the estimated effects of agglomeration are only marginally affected by the choice of units. At a given spatial scale, economic, administrative, and purely arbitrary units of more or less the same sizes imply roughly similar estimates. Consistent with the attenuation result reported above, there are some differences across spatial scales (when using, say, local labour markets vs regions), although the magnitudes remain close. Interestingly, the estimates are instead more sensitive to the chosen specification. For instance, controlling properly for the skills of the workforce (or not) can have a dramatic effect on the measured benefits of clustering (see below for more on this).

In conclusion, cluster proponents argue that clusters are extremely hard to measure both in terms of industry and geography. Even though it is true that clusters do not neatly coincide with industrial and spatial categories, existing empirical work suggests that the measurement errors made when assessing the benefits of clustering are small when using fine industrial definitions and metropolitan and regional spatial units.[40]

4.2 Cluster Productivity Benefits are Small and Uncorrelated with Local Wealth

A large literature attempts to measure various types of agglomeration effects (see Rosenthal and Strange, 2004, for a comprehensive survey). A sharp distinction is often made between what is known as 'localisation' (i.e., within sectors) and 'urbanisation' (i.e., across sectors) effects.

[39]Overall these findings suggest the existence of a 'two-tier clustering' phenomenon with 'narrow clusters' (narrowly defined industries at small spatial scales) surrounded by 'broad clusters' (related industries at a larger spatial scale). Given the short distances involved in narrow clustering, the latter is likely to be driven by direct interaction between workers (spillovers) and local labour market effects. The broader spatial scale at which vertically-linked industries co-localise suggests a more important role for trade linkages to explain broad clustering. This two-tier clustering raises a number of interesting issues regarding cluster policies which seem to be overlooked by cluster proponents.

[40]The interested reader can also refer to the more complete discussion of these issues in Rosenthal and Strange (2004).

Despite the claims of cluster proponents and in light of the discussion above, cluster effects are essentially equivalent to localisation effects (or, more precisely, we expect cluster effects to be captured through the estimation of localisation effects).

To assess the effects of clusters, the proper thought experiment is to ask *what would happen to a local outcome should an activity somewhere become larger, all else equal*. For instance, what is the effect of a larger local pharmaceutical industry on the productivity of its workers? Answering this type of question involves regressing a measure of individual productivity (such as the wage or the output per worker) on the local specialisation of the industry of employment, keeping total employment constant.

The large literature on localisation effects is supportive of the existence of benefits from clustering. In other words, the existing evidence is consistent with the idea that the productivity curve used in the above theoretical framework is upward sloping.

The first confirmations for the existence of clustering effects date back to Shefer (1973) and Sveikauskas (1975). For a subset of cities and sectors, Shefer (1973) found that a doubling of local employment in a sector would increase its productivity by up to 25%. Using a more robust approach, Sveikauskas (1975) found much smaller numbers, around 6 to 7%. The subsequent 20 years of research have found numbers very much in line with those of Sveikauskas (1975). More generally, the range of estimates for the mean elasticity of labour productivity to local industry employment is between 2 and 10% with a mid-point around 4 or 5% (Rosenthal and Strange, 2004).[41] Put differently, doubling the specialisation in a typical activity and area is associated with an increase in productivity of around 4%. This number is lower in some industries and higher in others (although it seldom goes above 10%). Even for industries with strong clustering effects, say, an elasticity of 8%, it takes more than a trebling of the specialisation of the cluster to see labour productivity increase by 10%. In conclusion, there are positive effects of clustering but the literature is also strongly suggestive that it takes extremely large increases in specialisation to get more than marginal effects on local productivity and wages.

Even though the effects estimated by the literature are very modest, they probably exaggerate the true (i.e., causal) benefits of clustering on productivity. First, many studies do not estimate a 'pure' effect of clustering (i.e., an increase in specialisation keeping total employment constant). Instead, these studies assess the effect of an increase in the size of the industry of employment, keeping employment in all other industries constant. Doing so conflates the effect of increased clustering with the effect of being in an area with more employment overall (since increasing the size of one industry keeping all other industries equal means an overall increase in employment). The second effect is arguably an 'urban growth' effect (or urbanisation effect) and not a cluster effect. Since this growth effect may be independent of the sector of employment, it should not be

[41]For the sake of brevity only the coefficient on local employment is discussed. Many studies try to capture the effects of clustering through the local number of firms and other measures of local industry 'size' as well as employment. Doing this does not affect the conclusions reported here.

counted as part of clustering. This (arcane) distinction matters a lot here because urbanisation effects are important. Cluster proponents do not argue in favour of growth in general but only for the development of specific activities.

Furthermore, most studies in the literature fail to control for the fact that causation may not run from clustering to high local productivity and wages but instead from the latter to the former.[42] If causation were to go in the opposite direction, i.e., from high productivity to high employment, all the results in the literature prior to the mid-1990s would be biased upward and thus exaggerate the magnitude of clustering effects. The first analysis that tackled this reverse-causation issue heads-on is due to Ciccone and Hall (1996). Determining the direction of causation is one of the hardest endeavours in applied economics. To do this, one needs to find some exogenous determinants of the variable that is suspected of endogeneity (e.g., local employment). The key restriction is that these exogenous determinants (or instruments) should be correlated with the variable to be explained (e.g., local productivity or local wages) only through the endogenous variable. Finding such determinants of local employment that are otherwise uncorrelated with local wealth is hard. For the US, Ciccone and Hall (1996) propose a number of them, which are related to the patterns of settlement of that country. Among others, they use the presence of early railroads in a county or its distance to the Eastern seaboard. They found that there was no large endogeneity bias when estimating agglomeration effects at the level of US states (wherein counties are aggregated in a particular way).

Using different sets of instruments and conducting their analysis at the level of smaller areas (French local labour markets instead of US states), Combes et al. (2008) and Combes et al. (2010) find evidence of some reverse causation. These findings are consistent with those of Ciccone and Hall (1996) since reverse causation is much more likely at a small spatial scale. This is because short-distance moves (or even adjustments in commuting patterns) are less costly than long-distance moves. Overall, the prevailing conclusion at this stage is that reverse causation probably biases existing estimates upwards, but only mildly so (Rosenthal and Strange, 2004; Combes et al., 2010).

A last concern with respect to the proper estimation of localisation effects is that the skills of workers need to be properly controlled for. The estimates are biased upwards again when better workers in the industry work in larger clusters. Arguably the most productive workers in the UK finance industry will be employed in London, its largest cluster. Those working in secondary financial centres such as Leeds are likely to be less productive. Finally those working where the industry is very small are likely to be even less productive, etc. Failing to control properly for skills will lead to the wrong attribution of the effects of greater skills to cluster size and thus bias the results. Controlling properly for skills is hard because many dimensions of skills

[42] Even when the mobility of the workforce is very low, substantial movements of workers can occur in the long-run. Going back to an example used above, the Californian wine production cluster may be bigger than that of Upstate New York because the returns to wine-making are intrinsically higher in California.

are unobserved by the statistician. Using the longitudinal dimension of their data for French workers, Combes et al. (2008) find this bias to be relatively large. The estimated elasticity of wages to specialisation drops from a cross-sector average of 4.3% to 2.1% when controlling for observed and unobserved skills. This suggests that the standard estimates of specialisation elasticities in the literature may need to be divided by two. If these low numbers were confirmed, cluster specialisation would need to become very high to obtain a non-trivial increase in labour productivity.

Even more negative conclusions are obtained by de Blasio and Di Addario (2005). Their study is of particular interest because they perform their analysis on Italian spatial units that were explicitly defined using a cluster mapping project. They do not find any positive evidence of a wage premium for workers in clusters. Their findings are even suggestive of lower returns to education in clusters! These results may be specific to Italian clusters (which would be ironic given the attention that Italian clusters have received in the literature). However, this lack of cluster effect is more likely to be caused by a subtle but interesting methodological difference. In their work, de Blasio and Di Addario (2005) attempt to assess the wage effect of working in a cluster regardless of the industry of employment whereas the rest of the literature looks at the effect of an increase in specialisation on the wages of workers in the same industry. There may be no contradiction between the two sets of results. The literature finds positive cluster effects for workers employed in the main sector of the cluster, while de Blasio and Di Addario (2005) fail to find evidence of broader benefits that would accrue to all workers.

At this stage, it is fair to conclude that there are small positive effects of clustering on labour productivity and wages. They appear to benefit only workers in the same sector of employment. There is still an open debate about whether these effects are small or very small, but the evidence presented here is certainly not supportive of large positive economic returns associated with cluster development. Nonetheless, the fact that *returns* to clustering are small may not run against the idea that clustering can make an important *contribution* to local productivity and local wages. It is not because a coefficient is small that the contribution to an outcome associated with its variable is necessarily small. Unfortunately, assessing the contribution of clustering to high local wages or high local productivity is not an exercise that is often performed in the economic literature.

Combes et al. (2008) are an exception. They report detailed results regarding the contribution of a range of local industry variables to individual wages. Their first result is that even when they are considered together, clustering variables (specialisation, number of local industry establishments, local share of industry professionals, etc) only make a very small contribution to local wages. More formally, the partial R^2 of the localisation variables in regressions purporting to explain individual wages are very low. This partial R^2 is, for instance, much smaller than that of other variables included in the regression such as labour market experience. In conclusion, the effects of clustering on wages and labour productivity are small and the overall contribution

of clustering to wages and productivity is even smaller.

Maybe much larger positive effects of clustering are to be seen elsewhere. The existing literature has used a broad range of variables to look at clustering effects. While the standard dependent variables in clustering regressions are output per worker, firm productivity, and wages, the literature has also considered firm creation, employment growth, or land rents. All these variables attempt to get at the productivity effects of clustering indirectly. This large set of papers (summarised in Rosenthal and Strange, 2004) draws conclusions that are very much in line with the results reported above. Clustering leads to better outcomes that are related to productivity but the effects are relatively small and cannot be used as justification for large scale local policies.

Beyond productivity, it may be that clustering affects a number of very different outcomes such as the intensity of labour market participation or innovation. About the former, Rosenthal and Strange (2008) show that workers in professional occupations, and especially the young ones, work more hours in larger cities while workers in other occupations tend to work fewer hours than in smaller cities. Among those other outcomes, innovation can be singled out because it matters a lot in the long run and also because it is very popular among cluster proponents (e.g., Porter, 2000a, 2003).[43] The literature on agglomeration effects with respect to innovation is much smaller than that looking at productivity outcomes. Nonetheless, the findings are generally not supportive of strong clustering effects on innovation (see Duranton and Puga, 2000, for a review). On the contrary, Feldman and Audretsch (1999), in a leading study, report that local innovation is fostered by a more diverse production structure. Local specialisation has instead a sizeable negative effect on innovation. Adding to this, the 'growth in cities' literature is generally supportive of a positive relationship between a diverse local production structure and subsequent employment growth (see Glaeser et al., 1992, and the subsequent literature).

In conclusion, clustering offers small 'static' productivity benefits and there is no strong evidence of positive dynamic (or innovation) benefits. Put differently the literature that attempts to assess the effects of clustering only offers very weak support to the claims made by cluster proponents.

All the studies examined so far assess the effects of clustering when it occurs 'naturally'. They do not constitute a direct assessment of cluster initiatives. Because cluster policies are very often local initiatives, it is very difficult to go beyond case studies and provide a reliable comparative exercise about their benefits. Work by Martin et al. (2011) is an exception. They examine a policy implemented by the French government in 1999. 'Local Productive Systems' offered subsidies to firms in particular sectors and particular areas to promote the development of dynamic clusters. Martin et al. (2011) show that the relative total factor productivity of

[43]This fascination with innovation may not be warranted here. At the aggregate level, a good argument can be made that innovation and welfare are strongly linked. The link between local innovation and local welfare could possibly be much more tenuous and should be scrutinised further.

treated firms over 1996-2004 declined by about 5%. Given the limited scope of this policy, it is difficult to imagine it could have such large negative impact. Instead, these results suggest that losers were selected or self-selected into this policy. This behaviour is consistent with the political-economy ideas developed above.

Before deepening the policy discussion, it is also important to understand where those clustering and agglomeration effects are coming from. Put differently, through which channel does clustering improve productivity? Knowing the answer to this would obviously be of prime importance for the proper design of cluster policy. Unfortunately, the literature on this issue does not provide very strong conclusions. Identifying the channels through which clustering effects percolate is extremely difficult. The reason behind this is that many clustering mechanisms share the same predictions regarding productivity outcomes.

A first solution is to use proxies for the various sources of benefits from clustering. Building on Dumais et al. (1997), this is what Rigby and Essletzbichler (2002) do. Their results show that local labour pooling plays a prominent role together with input-output linkages in determining agglomeration benefits. This is consistent with the more indirect findings of Dumais et al. (1997) and Rosenthal and Strange (2001). Nonetheless, some care is needed in the interpretation of these results due to significant econometric and data challenges. First, a stronger association between clustering benefits and an 'adequate' local labour market is obviously consistent with a causal link running from the latter to the former. However, the composition of the local labour market and high cluster benefits could be simultaneously determined. For instance, Silicon Valley's labour market may contain the exact right mix of workers but this may be a consequence (and not a cause) of a very successful cluster with costs of living so high that only those who stand to gain the most stay there. Second, the proxies used to measure labour market pooling may be much better than those used to measure technological spillovers (for which the paper trails are extremely scarce). Since those proxies are likely to be highly correlated, most of the correlation with clustering benefits may be sucked up by the better proxies regardless of the role that they actually play in clustering.

For some particular mechanisms research has been able to circumvent these data limitations and identification issues by focusing very narrowly on one specific mechanisms and ignore the others. For instance, Jaffe et al. (1993) attempt to get at the issue of local spillovers by examining geographical biases in patent citations. They ask whether patents developed nearby have a greater chance of being cited than patents developed further away, all else equal. Their initial results hinted at a substantial bias whereby patents originating from the same metropolitan areas are nearly three times as likely as being cited as patents from other metropolitan areas. Despite a very high level of care in the data work, subsequent research has disputed these findings and the current wisdom is that the local bias in citations is possibly no higher than 25% and short-lived as knowledge may diffuse nationally quite rapidly (Thompson, 2006). In a very different vein, Holmes (1999) shows that clustered firms are more likely to out-source and buy

their inputs locally, etc.

This type of work is important because it explores specific mechanisms and unique predictions of particular clustering stories. However, it only provides an 'existence proof' about these mechanisms and no magnitude regarding their importance.[44] Recent work has tried to map this type of approach into a productivity framework. Using workplace communication data at the level of individual workers, Charlot and Duranton (2004) find that up to 20% of the benefits from clustering percolate through workplace communication. Given the difficulty of finding the right data and the econometric challenges associated with this type of exercise, progress on the identification of the sources of clustering benefits can only be slow and tentative. Unfortunately detailed cluster policies will require knowing what matters to determine clustering benefits and where the inefficiencies are.

5 If not Clusters, What Should Local Economic Development Policies Be About Then?

The theoretical framework presented earlier was used to highlight the inadequacies of cluster policies. This section shows which 'positive' policy lessons may be learnt from it. Since this model is very crude, these policy recommendations are very sketchy and do not constitute a detailed blueprint for local economic development. They, nevertheless, provide a broad framework within which local development policies can be examined.[45]

5.1 "Pay More Attention to the Cost Curve than the Productivity Curve"

In relation to Figure 2, cluster policies appear to focus mostly on the productivity curve and neglect the complications associated with the cost curve and supply curve below. In this respect, cluster initiatives are not very different from a number of other policy recommendations for local development that have surfaced in the last 20 years such as fostering "learning regions" (Morgan, 1997) or attracting the "high bohemians" of the creative class (Florida, 2002), etc.

This near-exclusive focus on the productivity curve is misplaced. In the case of cluster proponents, this is because the policies are just too hard to design and implement, while their possible gains are elusive. Moreover, a good case can be made that productivity policies are probably best left to central governments.[46] Instead, larger gains can be achieved locally by

[44]There is also a large literature in geography that documents the workings of clusters at a great level of detail. This tradition was pioneered in series of studies edited in Pyke et al. (1990). The work of Saxenian (1994) on Silicon Valley is another classic. This literature is reviewed in Storper (1997), Maskell (2001), and Martin and Sunley (2003). Markusen (1996) and Gordon and McCann (2000) also provide interesting generalisations.

[45]See Combes et al. (2005) for more on this.

[46]A higher productivity often materialises into lower prices which benefit all consumers and while local producers may not benefit much. Hence local policies may not be a good idea when the benefits are so broadly distributed. Besides, the market failures that can justify those productivity policies take place

improving the policies associated with the cost curve, such as land-use planning, urban transport, provision of local public goods, etc. These policies, which have been for the longest time the 'staple' policies of local governments, may not be as 'sexy' as setting up a bio-tech cluster or the next Silicon Valley. They are nonetheless fundamental determinants of how easy it is to produce in a place and how pleasant it is to live there.

Still, good policies associated with the cost curve are difficult to achieve.[47] For instance, urban transport policies raise many tough questions. What should downtown parking provision be like? Should there be a congestion tax? Should the supply of taxis be regulated? If yes, how? Should new roads be built? Should the public transit network be extended? How should it be priced? Are the latest traffic management technologies worth the investment? All these questions (and many more) are hard to answer. Despite all this, a good transport policy is easier to implement than a good cluster policy.[48] This is because we know a lot more about the inefficiencies of urban traffic than those taking place in local production. Furthermore, a bad transport policy is often for all to observe unlike a bad cluster policy.

Finally, when they try to develop clusters and interfere with the local composition of economic activity, local authorities are likely to overreach themselves. Given their limited capabilities, local government may spread themselves too thin by trying to do too much.[49] This may have a negative impact on their other policies. To conclude, the recommendation for local governments is thus to improve their traditional areas of intervention rather than try to do 'new things'.

5.2 "Go for Growth, Sometimes"

The second set of local policies has to do with the supply curve. Although some cities may be too big, in a world of upward-sloping supply curves as in Figure 4, most places may be too

---

everywhere so that a common (i.e., national) set of corrective policies may be desirable.

[47]One may also worry whether correcting inefficiencies on the cost side makes sense in light of the inefficiencies associated with the productivity curve. First, it is unlikely that fixing local congestion will make the other inefficiencies so much worse as to lower net returns. Using Figure 2, it is true that one local authority in isolation will not improve net returns by lowering local costs. This is because any gain on the cost side will lead to the arrival of newcomers and net returns remain equal to $A_Z - C_Z$. This complete crowding out of the costs improvement is caused by perfect labour mobility. From Figure 4, lower costs lead to an increase in net returns in absence of perfect mobility. Besides, as exposed below, if many jurisdictions lower their costs, the net returns will improve in many places and this will lead to a higher supply curve and thus a better allocation in the economy.

[48]This does not prevent local government from getting their transport policies spectacularly wrong sometimes as Paris did recently (see Prud'homme et al., 2005).

[49]Focusing the effort on the cost curve does not imply that local authorities should always try to micromanage everything. A good case may be made that local authorities should be more involved in the details of traffic management and possibly less so in some of the details of land use.

small.[50] These raises two questions. Does 'going for growth' make sense? And if yes, how should it be done?

To answer the first question, being bigger does yield some economic benefits (despite the reluctance to admit it in some quarters). These benefits from size are often called 'urbanisation economies'. They can be measured using the same regression approach used to measure the benefits from clustering. There is a vigourous academic debate about the elasticity of local wages and local productivity with respect to total employment. The current consensus is that, in terms of elasticities, urbanisation effects are about as large as localisation effects (Rosenthal and Strange, 2004; Combes et al., 2008). Even though the elasticities are about the same, a larger size seems to be more desirable than increased specialisation for several reasons. First, the contribution of size to local productivity and wages is much larger than that of specialisation (Ciccone and Hall, 1996; Combes et al., 2008) with a much higher partial R^2 in the regressions. Then, it is arguably much easier to "go for growth" (i.e., increase local employment irrespective of sector) than expand employment in a particular activity. Finally, diversified growth is less risky than increased cluster specialisation.

However, going for growth is no panacea. In particular, there is a sharp tradeoff in some places between employment and population growth, on the one hand, and local amenities, on the other. Unsurprisingly, many places with outstanding amenities (from San Francisco to central Paris) have implemented very restrictive policies instead of going for further growth. Again, one may take issue with the way those restrictive policies are implemented and criticise their excessive restrictiveness. However, there is very little doubt however that central Paris would lose much of its valuable charm if development was allowed to proceed unfettered.

Should the case for growth be strong enough (and it appears to be quite strong in many places), how should growth be fostered? However 'policy incorrect' this may sound, the recent literature appears to indicate that the recipe for growth involves being attractive, sprawly, and opportunistic. Attractive means providing good local amenities from parks and safety to good primary schools. It is hard to assess the local growth effects of better amenities in general. Studies that concentrate on climate find that a nice weather is a first-order determinant of local growth (Rappaport, 2007). It is of course hard to affect local climate.[51] But other amenities can be improved by local policies.

There is also mounting evidence showing that cities exhibiting more of what is often referred to pejoratively as urban sprawl may enjoy better outcomes (Glaeser and Kahn, 2001 and 2004) and possibly faster growth (Baum-Snow, 2007). This should not come as a surprise. Cities that offer their residents the type of housing they like and easy commutes to jobs should fare better

[50] It might be argued that many Western European countries such as France, Italy, Spain, and the UK have too few big cities.

[51] The fact that climate is exogenous makes it easier to assess it as a determinant of local growth. Assessing the effect of other amenities is plagued by endogeneity problems.

than those that prevent newcomers from settling in and offer inconvenient commutes.[52] However, "be sprawly" does not mean that a free-for-all type of development is desirable. Recall that the first recommendation is to be attractive.[53] Most places need to achieve a balance between preservation (environmental or architectural) and affordable housing (and urban efficiency more generally). This balance will differ across places. However a good case can be made that too many places are too restrictive (see Glaeser et al., 2006, for the US and Cheshire and Sheppard, 2002, for the UK).

Back to the baseline model, the rationale for being attractive is that better amenities lower the supply curve. For sprawl, affordable housing and easy commutes raise the cost curve (i.e., they lower local costs). In light of this framework, the reasons for 'opportunism' are also easy to understand when the supply curve is less than perfectly elastic. Employment growth is hard to boost in 'normal conditions' since established firms are nearly immobile and start-ups take many years to develop (if ever). Newly created large establishments searching for a location represent quick opportunities for significant employment growth. Such growth opportunities do not come for free however because there are usually a number of jurisdictions competing for them. This brings us to the last recommendation.

5.3 "Central Governments, Let Them Compete"

The argument developed here against cluster policies could be viewed as part of a broader argument against territorial competition. The latter is usually regarded with much suspicion in the academic literature (see Cheshire and Gordon, 1998, for a review). The case against territorial competition, as made in the literature, is in two parts. First, territorial competition is accused of being at best a zero-sum game. Since the establishment needs to locate somewhere (so goes the argument), there is no social gain from territorial competition, only losses due to the cost of competing. Although it is true that the competition process may be costly (cost of learning about existing opportunities, bidding, etc), we also expect large new plants to generate positive local external effects (by the same logic exposed above for clustering). The crucial point is that these external effects can vary across places. In absence of territorial competition, plants may end up in the 'wrong' location.[54] With territorial competition, the places for which the external effects are the strongest are expected to bid the most. Hence, at the (modest) cost of the competition process, territorial competition can improve the spatial allocation of plants.

The second part of the traditional case against territorial competition stems from its (possibly

[52] This is consistent with the recommendation above regarding the focus on the cost curve.

[53] Although sprawliness and attractiveness may conflict at times, a city can manage both. Portland and Seattle score very high on the sprawl measure of Burchfield et al. (2006). These cities are nonetheless widely viewed as attractive. At the opposite end, Detroit is both compact and often deemed not very attractive.

[54] In absence of territorial competition, plants choose their location as to maximise their *private* profit, ignoring external effects.

large) redistributive element (from places to firms) with places paying 'too much' to get the plants. While there is certainly a theoretical possibility of a 'curse of the winner' associated with this type of rivalry, the social gain (or loss) of the winner is ultimately an empirical question.[55] This is very hard to investigate because of the very large number of confounding factors that can explain the evolution of the places that get these plants. In a very clever recent piece, Greenstone and Moretti (2004) gathered data about not only the winners of a large number of such contests, but also their 'runner-ups' (jurisdictions that also bided and came to close winning but ultimately lost). Since the runner-ups are arguably very similar *ex-ante* to the winners, *ex-post* differences are very likely to be mostly due to winning or losing such a contest. The comparison of the winners and runner-ups, before and after, is suggestive of sizeable local gains associated with the successful attraction of new plants.[56]

This argument in favour of territorial competition is actually broader than the narrow issue of bidding for new plants. To avoid the pitfalls of cluster policies, there is a temptation to restrict the powers of local governments with respect to local development or instead conduct cluster policies at the national level as in France. Both temptations should be resisted. National cluster policies are subject to much of the same problems as local cluster initiatives. Government failures may be even more serious with centralised cluster policies because of the political pressure to spread the subsidies across the country.[57] This runs against the strong concentration tendencies at the heart of the benefits from clustering. Despite the strong case developed here against cluster policies, preventing local governments from being in charge of their destinies would be wrong.[58]

In a large majority of jurisdictions in developed economies we expect homeowners to form a majority and elect local officials acting in their interest. In practice, this implies trying to maximise aggregate land values in their jurisdiction. In absence of strong interdependencies across jurisdictions, this local maximisation of land values can lead to efficient outcomes. In brief, we expect local characteristics such as good amenities or high wages to be capitalised into land values. Since land prices reflect how much residents and firms are willing to pay to be somewhere, maximising land values can be synonymous with the maximisation of the

[55]Notwithstanding the fact that these transfers may increase aggregate investment and improve its efficiency.

[56]These bids are not usually part of cluster strategies. They belong instead to broad-based policies that attempt to promote local growth or reverse local decline.

[57]The recent French cluster policy that subsidises more than 60 cluster efforts across the country is a case in point.

[58]Although, this piece is very critical of cluster policies and raises more general doubts about other types of local 'productivity' policies, it does not advocate curtailing the power of local authorities in matters of local economic development. For instance, local governments can act as a catalyst (and no more) between local producers and universities at very low cost. Despite the skepticism expressed here, it might also be that a local government will some day find a 'good formula' for local economic development that can then be applied elsewhere.

attractiveness of a place.[59]

We return to Figure 2. As local governments try to maximise their land values and the attractiveness of their own jurisdiction, they also shift the supply curve of other jurisdictions upwards. In turn, to avoid an exodus these other jurisdictions need to offer a 'better deal' to their residents. In practice, a better deal can mean a higher quality of life (through a lowering of costs and better amenities) or higher incomes and more economic opportunities (through a growth strategy). Then, and as can be seen from Figure 2, a higher supply curve reduces the inefficiency associated with the cluster co-ordination failure. Put differently, with the natural tendency of local economies to specialise (at least some them, as shown above), territorial competition should lead to more efficient clusters without any direct intervention into the production process by local authorities.

Many may disagree with this sketch (or think there is much to disagree with it). However, the main point to remember is that local government should focus on the issues for which they can make a valuable difference. This means a concentration on the policies associated with their cost curve rather than their productivity curve, and deciding whether they want to go for growth or preservation. For central governments, who should take a national approach to productivity related issues, the focus should be on the governance of their sub-national units and how the latter should compete.

6 Conclusions

Clusters are a complex second-order issue that wrongly receive first-order attention. More specifically, the case against clusters policies is in two parts. First, good cluster policies involve solving a very difficult co-ordination problem and correcting for a number of market failures, which we know very little about. Adding to the difficulty, cluster policies need to be designed and implemented in very uncertain environments without being captured by special interests. Second, even if the policy-makers can get it right, the benefits of clustering are simply too small empirically to justify significant and sustained efforts towards clusters. Californian clusters may be very prosperous but most clusters are not particularly so. Instead of dreaming about them local policy-makers should focus their attention away from the local production structure and aim instead at a more efficient provision of public goods that serve the needs of both residents and a broad range of local producers.

[59]This increase in land prices can then be taxed away to ensure an optimal provision of local public goods. It is beyond the scope of this report to enter the detail of this argument. See Becker and Henderson (2000) and Wildasin (2002) for recent complete treatments. These ideas date back to George (1884) and Tiebout (1956).

References

Arzaghi, M. and J.C. Davis (2005), Spatial Mobility and Geographic Concentration, mimeographed, US Census Bureau.

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

Baldwin, R.E. and F. Robert-Nicoud (2007), Entry and Asymmetric Lobbying: Why Governments Pick Losers, *Journal of the European Economic Association* 5, 1064–1093.

Baum-Snow, N. (2007), Did Highways Cause Suburbanization? *Quarterly Journal of Economics* 122, 775–805.

Becker, R. and J.V. Henderson (2000), Intra-Industry Specialization and Urban Development, in J.M. Huriot and J.-F. Thisse (eds.) *Economics of Cities: Theoretical Perspectives* Cambridge, Cambridge University Press.

Belleflamme, P., Picard, P. and J.-F. Thisse (2000), An Economic Theory of Regional Clusters, *Journal of Urban Economics* 48, 158–184.

Besley, T. (2006), *Principled Agents? The Political Economy of Good Government*, Oxford, Oxford University Press.

Briant, A., Combes, P.P., and M. Lafourcade (2010), Does the Size and Shape of Geographical Units Jeopardize Economic Geography Estimations? *Journal of Urban Economics* 67, 287–302.

Burchfield, M., Overman, H.G., Puga, P. and M.A. Turner (2006), Causes of Sprawl: A Portrait From Space, *Quarterly Journal of Economics* 121, 587–634.

Caves, R.E. (1998), Industrial Organization and New Findings on the Turnover and Mobility of Firms, *Journal of Economic Literature* 36, 1947–1982.

Charlot, S. and G. Duranton (2004), Communication Externalities in Cities, *Journal of Urban Economics* 56, 581–613.

Cheshire, P. and S. Sheppard (2002), The Welfare Economics of Land Use Planning, *Journal of Urban Economics* 52, 242–269.

Cheshire, P.C. and I.R. Gordon (1998), Territorial Competition: Some Lessons for Policy, *Annals of Regional Science* 32, 321–346.

Ciccone, A. and R.E. Hall (1996), Productivity and the Density of Economic Activity, *American Economic Review* 86, 54–70.

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

Combes, P.-P., Duranton, G., Gobillon, L. and S. Roux (2010), Estimating Agglomeration Economies with History, Geology, and Worker Effects, in E.L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge (MASS.), National Bureau of Economic Research.

Combes, P.-P., Duranton, G., Gobillon, L., Roux, S., and D. Puga (2009), The Productivity

Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection, mimeographed, University of Toronto.

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

Cortright, J. (2006) Making Sense of Clusters: Regional Competitiveness and Economic Development, Discussion paper, The Brookings Institution Metropolitan Policy Program.

Council on Competitiveness, Monitor Company, and M. Porter (2001), *Clusters of Innovation*, Washington: National Report.

de Blasio, G. and S. Di Addario (2005),Do Workers Benefit from Industrial Agglomeration? *Journal of Regional Science* 45, 797–827.

Department of the Environment, Transport and the Regions (2000), *Planning for Clusters: A Research Report*, London, DETR.

Department of Trade and Industry (2001a), *Business Clusters in the UK: A first Assessment*, London, DTI.

Department of Trade and Industry (2001b), *Raising UK Productivity – Developing the Evidence Base for Policy*, London, DTI.

Devereux, M.P., Griffith, R. and H. Simpson (2007), Agglomeration, Regional Grants and Firm Location, *Journal of Public Economics* 91, 413–435.

Dumais, G., Ellison, G. and E.L. Glaeser (1997), Geographic Concentration as a Dynamic Process, Working Paper 6270, National Bureau of Economic Research.

Duranton, G. (2007), Urban Evolutions: The Fast, the Slow, and the Still, *American Economic Review* 97, 197–221.

Duranton, G. and H.G. Overman (2005), Testing for Localization using Micro-Geographic Data, *Review of Economic Studies* 72, 1077–1106.

Duranton, G. and H.G. Overman (2008), Exploring Detailed Patterns of UK Manufacturing Location Using Microgeographic Data, *Journal of Regional Science* 48, 213–243.

Duranton, G. and D. Puga (2000), Diversity and Specialisation in Cities: Why, Where and When Does it Matter? *Urban Studies* 37, 533–555.

Duranton, G. and D. Puga (2001), Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products, *American Economic Review* 91, 1454–1477.

Duranton, G. and D. Puga (2004), Micro-Foundations of Urban Agglomeration Economies, in J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4, Amsterdam, North-Holland.

European Commission (2003), *Final Report of the Expert Group on Enterprise Clusters and Networks*, Brussels, European Commission.

Faberman, R.J. (2005), What's in a city? Understanding the Micro-Level Employer Dynamics Underlying Urban Growth, Working Paper 386, US Bureau of Labour Statistics.

Feldman, M.P. and D.B. Audretsch (1999), Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition, *European Economic Review* 43, 409–429.

Fischel, W.A. (2000) Zoning and Land Use Regulations, in B. Boudewijn and G. De Geest (eds.) *Encycolopedia of Law and Economics*, volume 2, Cheltenham, Edward Elgar.

Florida, R. (2002), *The Rise of the Creative Class. And How It's Transforming Work, Leisure and Everyday Life*, New York, Basic Books.

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

George, H. (1884) *Progress and Poverty: An Inquiry into Causes of Industrial Depressions, and of Increase of Want with Increase of Wealth. The Remedy*, London, W. Reeves.

Glaeser, E.L. (2005), Reinventing Boston: 1630-2003, *Journal of Economic Geography* 5, 119–153.

Glaeser, E.L., Gyourko, J. and R.E. Saks (2006), Urban Growth and Housing Supply, *Journal of Economic Geography* 6, 71–89.

Glaeser, E.L. and M. Kahn (2001), Decentralized Employment and the Transformation of the American City, *Brookings-Wharton Papers on Urban Affairs*, 1–47.

Glaeser, E.L. and M.E. Kahn (2004), Sprawl and Urban Growth, in J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4, Amsterdam, North-Holland.

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

Gordon, I.R. and P. McCann (2000). Industrial Clusters: Complexes, Agglomeration and/or Social Networks? *Urban Studies* 37, 513–532.

Greenstone, M. and E. Moretti (2004), Bidding for Industrial Plants: Does Winning a 'Million Dollar Plant' Increase Welfare? mimeographed, University of California Berkeley.

Greenwood, M.J. (1997), Internal Migrations in Developed Countries, in M.R. Rosenzweig and O. Stark (eds.) *Handbook of Population and Family Economics*, volume 1B, Amsterdam, North-Holland.

Henderson, J.V. (1974), The Sizes and Types of Cities, *American Economic Review* 64, 640–656.

Holmes, T.J. (1999), Localisation of Industry and Vertical Disintegration, *Review of Economics and Statistics* 81, 314–325.

Jaffe, A.B., Trajtenberg, M. and R. Henderson 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics* 108, 577–598.

Leslie, S.W. and R.H. Kargon (1996), Selling Silicon Valley: Frederick Terman's Model for Regional Advantage, *Business History Review* 70, 435–482.

Markusen, A.R. (1996), Sticky Places in Slippery Space: A Typology of Industrial Districts, *Economic Geography* 72, 294–314.

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

Martin, P. (1999), Public Policies, Regional Inequalities and Growth, *Journal of Public Economics* 73, 85–105.

Martin, P., Mayer, T. and F. Mayneris (2011), Public Support to Clusters: A Firm Level Study of French 'Local Productive Systems', *Regional Science and Urban Economics* 41, 108–123.

Martin, R. and P. Sunley (2003), Deconstructing Clusters: Chaotic Concept or Policy Panacea, *Journal of Economic Geography* 3, 5–35.

Maskell, P. (2001),Towards a Knowledge-Based Theory of the Geographical Cluster, *Industrial and Corporate Change* 10, 921–943.

Morgan, K. (1997), The Learning Region: Institutions, Innovation and Regional Renewal, *Regional Studies* 31, 491–503.

Obstfeld, M. and G. Peri (1998), Regional Non-Adjustment and Fiscal Policy, *Economic Policy* 13, 640–656.

OECD. 2001. *Innovative Clusters: Drivers of National Innovation Systems*, Paris, OECD.

Ottaviano, G.I.P. (2003), Regional Policy in the Global Economy: Insights From the New Economic Geography, *Regional Studies* 37, 665–673.

Pellenbarg, P.H. (2005), Firm Migration in the Netherlands, mimeographed, University of Groningen.

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

Porter, M.E. (1998), Clusters and the New Economics of Competition, *Harvard Business Review* 76, 77–91.

Porter, M.E. (2000a), Location, Competition, and Economic Development: Local Clusters in a Global Economy, *Economic Development Quarterly* 14, 15–34.

Porter, M.E. (2000b), Locations, Clusters, and Company Strategy, in G.L. Clark, M.P. Feldman, and M.S. Gertler (eds.) *The Oxford Handbook of Economic Geography*, New York, Oxford University Press.

Porter, M.E. (2003), The Economic Performance of Regions, *Regional Studies* 37, 549–578.

Prud'homme, R., P. Kopp, and J.P. Bocarejo (2005), Evaluation de la Politique Parisienne des Transports, mimeographed, University of Paris I.

Puga, D. (2002), European Regional Policy in Light of Recent Location Theories, *Journal of Economic Geography* 2, 372–406.

Pyke, F., G. Beccattini, and W. Sengenberger (1990), *Industrial Districts and Inter-firm Cooperation in Italy*, Geneva, International Institute for Labour Studies.

Rappaport, J. (2007), Moving to Nice Weather. *Regional Science and Urban Economics* 37, 375–398.

Rigby, D.L. and J. Essletzbichler (2002), Agglomeration Economies and Productivity Differences in US Cities, *Journal of Economic Geography* 2, 407–432.

Rosenthal, S.S. and W.C. Strange (2001), The Determinants of Agglomeration, *Journal of Urban Economics* 50, 191–229.

Rosenthal, S.S. and W.C. Strange (2003), Geography, Industrial Organization, and Agglomeration, *Review of Economics and Statistics* 85, 377–393.

Rosenthal, S.S. and W.C. Strange (2004), Evidence on the Nature and Sources of Agglomeration Economies, in J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4, Amsterdam, North-Holland.

Rosenthal, S.S. and W.C. Strange (2008), Agglomeration and Hours Worked, *Review of Economics and Statistics* 90, 105–118.

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

Shefer, D. (1973), Localization Economies in SMAs: A Production Function Analysis, *Journal of Regional Science* 13, 55–64.

Solvell, O., Lindqvist, G. and C. Ketels (2003), *The Cluster Initiative Greenbook*, Gotheburg, Competitiveness Institute (TCI)/Vinnova.

Storper, M. (1997), *The Regional World: Territorial Development in a Global Economy*, New York, Guilford.

Sveikauskas, L.A. (1975), The Productivity of Cities. *Quarterly Journal of Economics* 89, 393–413.

Thompson, P. (2006), Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-Added Citations, *Review of Economics and Statistics* 88, 383–389.

Tiebout, C.M. (1956), A Pure Theory of Local Expenditures, *Journal of Political Economy* 64, 416–424.

van der Linde, C. (2003), The Demography of Clusters – Findings From the Cluster Meta-Study, in J. Bröcker, D. Dohse, and R. Soltwedel (eds.) *Innovation Clusters and Interregional Competition*, Berlin, Springer Verlag.

Wildasin, D. (2002) *Urban Public Finance*. London, Harwood Academic Publishers.