Agglomeration and clusters: Tools and insights from coagglomeration patterns[*]

Kristian Behrens[†]

September 23, 2016

Abstract: Geographic clusters have received first-order policy attention despite our limited academic knowledge about them. I explain how coagglomeration patterns of industries can inform the analysis of clusters by allowing us to better delineate them; to explore the causal mechanisms underlying them; and to more finely assess their impact on various economic outcomes.

Keywords: Geographic clusters; coagglomeration; agglomeration mechanisms; economic policy.

JEL Codes: R11; R12; R58.

---

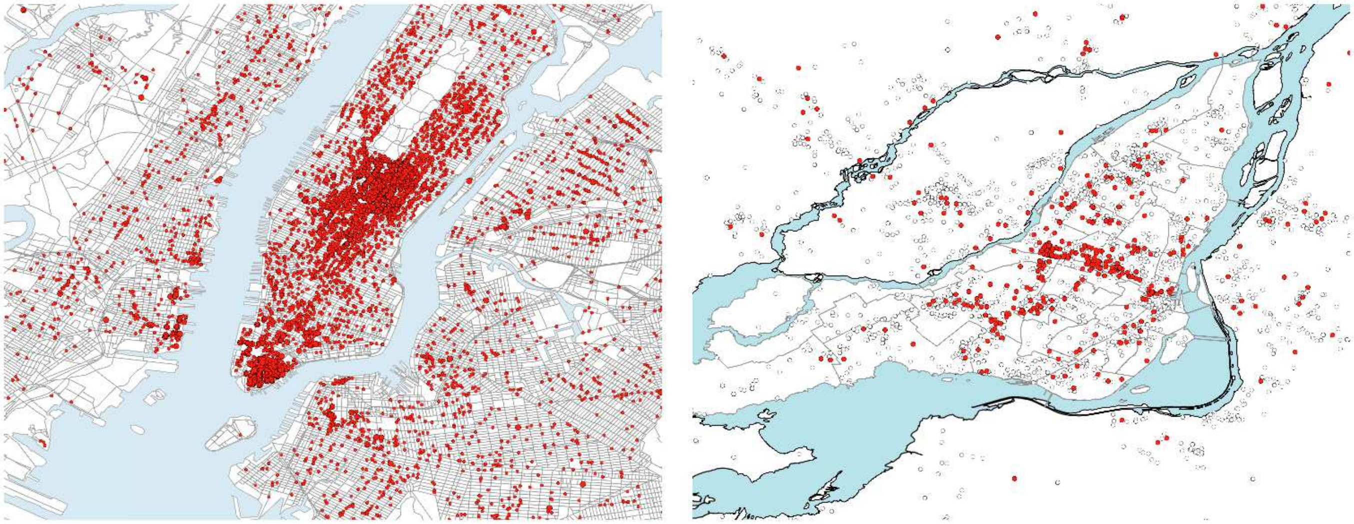
[*] I thank Yves Duclos and Nancy Gallini for giving me the opportunity to deliver the 2016 Innis Lecture. This lecture largely grew out of joint work with Brahim Boualam, Théophile Bougna, Mark Brown, Gilles Duranton, Rachel Guillain, Julien Martin, Florian Mayneris, Giordano Mion, Yasusada Murata, and Frédéric Robert-Nicoud, all of whom contributed in various ways to my understanding of the different aspects of the problems at hand. None of them is to blame for any remaining errors, imprecisions, and omissions, which are solely mine. I thank the co-editor in charge, Dan Bernhardt, as well as Rachel Guillain, Florian Mayneris, and especially Will Strange for valuable comments on an earlier draft of this paper. I gratefully acknowledge financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada for the funding of the *Canada Research Chair in Regional Impacts of Globalization*. This study has also been funded by the Russian Academic Excellence Project '5-100'.

[†] École des Sciences de Gestion, Université du Québec à Montréal, Canada; National Research University Higher School of Economics, Russian Federation; CIRPÉE; and CEPR. E-mail: `behrens.kristian@uqam.ca`

1. Introduction

The geographic concentration of economic activity is a first-order feature of the data. Its most visible manifestation are large cities: in 2001, the six largest Canadian metropolitan areas were home to 45.8% of the population, on barely 0.35% of the country's surface. This geographic concentration takes various forms at different spatial scales, ranging from broad 'manufacturing belts' — like the Québec City-Windsor corridor — to more localized industrial concentrations like the automotive industries in the Detroit-Oshawa portion of that corridor. Clustering also takes place even more locally at smaller scales, as can be seen from our cities' central business districts or the concentration of certain types of shops along specific streets or in shopping malls.

Figure 1: Geographic concentration is widespread across industries.



(a) 'Wall Street', NYC.

(b) 'Boulevard Saint-Laurent', MTL.

Panel (a) of Figure 1 depicts the clustering of the 'Finance and Insurance' industry (NAICS 52) in New York City, essentially Manhattan. While most people would have no difficulty to recognize that figure, panel (b) would pose more problems. It depicts the location of the 'Textile and Clothing' industries (NAICS 3131–3169) — the so-called 'Shmata' — in Montréal, mostly along Saint-Laurent boulevard.[1] Figure 1 highlights that although geographic concentration is often identified with its emblematic examples — Silicon Valley, the North Carolina Research Triangle, or Wall Street — it is far more ubiquitous and not limited to high-tech, bio-tech, or finance. The carpet industry in Dalton, Georgia (Krugman, 1991), the furniture industry in

---

[1] "Le boulevard Saint-Laurent a été pendant près de soixante ans le centre de la confection de vêtements au Canada." (Pierre Anctil, available online at `http://www.atsa.qc.ca/ressources/fichiers/frags_2/downloads/10FRAGS_Rallye _r.pdf`). Though its heydays are long gone, the historical heritage of the textile and clothing industry still makes for some of the most strongly concentrated industrial clusters in Canada in the first decade of the 2000s (see Behrens and Bougna, 2015; and Behrens, Boualam, and Martin, 2016a,b).

Raleigh, North Carolina (Strange, 2009), or 'Boulevard Saint-Laurent' in Montréal are just as real examples of clusters than Wall Street or 'the Valley'.

Geographic concentration of economic activity can occur for a variety of reasons. It reflects, at least partly, the 'locational fundamentals' of Harold Innis' writings: some places offer specific advantages to some activities, which thus end up concentrating there.[2] Less obvious, geographic concentration also reflects the interrelatedness of economic agents along various dimensions. This interrelatedness — which can take many forms such as buyer-supplier relationships, intended or unintended knowledge exchanges, or labor mobility across firms and industries — enhances productivity, especially when agents are in close geographic proximity. The consensus estimates for the elasticity of productivity to geographic concentration — known in the literature as *agglomeration economies* — lie somewhere between 1.5%–5%, with recent estimates using better approaches and microdata being at the lower end of that range. The empirical evidence further strongly suggests that the relationship from geographic concentration to productivity is causal.[3]

The existence of agglomeration economies is the basic driving force for the formation of *clusters*, defined as "a geographically proximate group of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities." (Porter, 2000, p.16). Clusters tend to form naturally because geographically mobile profit-maximizing agents seek to take advantage of the productivity gains generated by the spatial concentration of related economic activities. If this was the end of the story, there would be no need for policy since the market would deliver clusters of optimal size and composition. Yet, most of the mechanisms that generate productivity gains in clusters involve some form of market failure. For example, uncompensated externalities from knowledge spillovers or congestion in local markets may lead to either a suboptimal level of geographic concentration, or the 'wrong mix' of industries in some locations, or both. More fundamentally, mobile agents disregard numerous effects they have on the localities they leave or enter, and the absence of coordination in individual migration choices makes the emergence of efficient clusters unlikely and may harm those 'left behind'. The existence of productivity gains from clustering, when combined with the numerous market failures that arise in clusters and during the process of

---

[2] Harold Innis wrote mostly on Canadian economic history, looking in particular at natural resources and locational fundamentals ("Problems of Staple Production in Canada"; "Settlement and the Mining Frontier"; "The Cod Fisheries"), transportation ("A History of the Canadian Pacific Railway"), and trade ("The Fur Trade in Canada"). Some of Innis' work influenced the way people think about regional economies. As mentioned by Robin Neill, Innis' "staple, primary products or export base theory of economic development, was extended by Douglass North in applications to regional development in the United States [...] subsequently it was elaborated in generalized export-base models [...]." (available online at `http://eh.net/encyclopedia/harold-adams-innis/`). Export-base theory and local multipliers remain staples (no pun intended) of regional analysis.

[3] See Rosenthal and Strange (2004), and Combes and Gobillon (2015), for surveys of the empirics of agglomeration; and Duranton and Puga (2004), and Behrens and Robert-Nicoud (2015), for surveys of the theory. Combes, Duranton, and Gobillon (2011) provide a good discussion of the causality question.

their formation and growth, are the starting point for, and the underlying rationale of, different types of cluster policy.

Cluster policy has become very fashionable. There have been numerous studies and initiatives on clusters during the first decade of the years 2000 in Canada.[4] There are also several provisions in the Federal Budget 2016 that suggest a federal agenda to the cluster question.[5] Clusters and cluster policy are, therefore, still very much in the spotlight. Observe that this hype is not specific to the Canadian context. A simple search on Google Scholar yielded 1.43 million hits for the two keywords 'economic' and 'clusters' as of August 2016. Clusters, therefore, have also attracted interest in many other countries that have developed cluster policies and run various 'cluster programs', 'cluster initiatives', or 'cluster observatories'.[6]

There are three main reasons for the broad acceptance of cluster policy in the policy arena. The first is linked to the real or perceived benefits of clusters. The starting premise of cluster policy — that clusters are associated with productivity gains — seems correct. When combined with the existence of externalities, this is a strong argument in favor of cluster policy. However, most academic economic research fails to find strong effects of cluster policy. The reason is that the productivity gains appear small and are, as explained above, already largely internalized by firms' location decisions. It is also unclear that policy can affect the geographic concentration of industries sufficiently for the potential gains to outweigh the costs (see Duranton et al., 2010; Martin, Mayer, and Mayneris, 2010). Geographic concentration also tends to get associated in people's minds with all sorts of other positive outcomes: faster innovation, more human capital formation or, more generally, vague positive ideas such as regional 'prosperity' or 'competitiveness'. We have little evidence to back these associations, and even less evidence suggesting that causal mechanisms are at work. As squarely summarized by Duranton (2011, p.3): "[the case] for cluster policy [. . .] is theoretically ambiguous and empirically very weak." The second reason is linked to the strong marketing of the 'cluster brand', which has spawned a whole industry that has been very succesful at selling cluster policies to policy makers (see

---

[4] Without being exhaustive, let me mention the Innovation Systems Research Network (ISRN) study, 2000-2005 (see, e.g. Wolfe and Gertler, 2004); the Conference Board of Canada (2004) study "Clusters of Opportunity, Clusters of Risk"; and Canada's Innovation Strategy 2002, that had the target of developing at least ten internationally recognized technology clusters by 2010. Trefler (2008) noted in his 2008 Innis Lecture that the spatial organization of economic activity ('cluster mix', 'cluster effectiveness', and 'urbanization') has a strong bearing on the U.S.-Canada prosperity gap, equal in magnitude to the contribution of both 'education' and 'capital investment'. Strange (2009) also surveyed 'agglomeration research in the age of disaggregation' in his 2009 CEA state-of-the-art lecture. More recently, the Institute for Competitiveness and Prosperity in Toronto has developed the Canadian Cluster Data project.

[5] Budget 2016, Chapter 2 (available online at `http://www.budget.gc.ca/2016/docs/plan/ch2-en.html`) states: "Budget 2016 proposes to make available up to $800 million over four years, starting in 2017–18, to support innovation networks and clusters as part of the Government's upcoming Innovation Agenda." Furthermore, "Budget 2016 announces the Government's intent to develop, in collaboration with provinces, territories, research institutions and other stakeholders, a nationwide Canadian Cluster Mapping portal."

[6] Examples include the 'Kompetenznetze' in Germany; the 'Pôles de competitivité' in France: the 'Cluster Mapping Project' in the U.S.; and the 'National Cluster Program' in Russia.

Martin and Sunley, 2003, for a critical discussion). As I discuss below, this type of cluster policy is rarely backed by solid empirical evidence and sound theories. Last, in a world where more traditional industrial policies have either fallen into discredit or are no longer applicable — due to various agreements that prohibit explicit subsidies to industries— cluster policy is increasingly seen as one of the few remaining first-order policy levers, especially at the local level (Duranton, Martin, Mayer, and Mayneris, 2010). Since 'the local' has become increasingly more important to voters and policy makers alike, this partly explains the lure of those policies.

Given the strong policy interest in clusters, we would expect that there are solid theoretical and empirical bases to inform the debates and to back important policy decisions. However, the truth is that while academic work — pushed in particular by geographers but also by regional and urban economists — has become increasingly interested in the question of geographic concentration, there are still major gaps in what we know about clusters. Clusters represent one area in which the distance between scientific knowledge and policy demands has not been narrowing substantially. We do not know enough about clusters to make sound policy recommendations. This situation creates a dangerous void where large sums of money are gambled on simple and generic advice that is rarely backed by more systematic knowledge.[7] There is an urgent need to go beyond 'commented experiences of good and less good practice' when it comes to cluster policy. Cluster policy requires at least some statistical inference and external validity, and it should hold up to closer scrutiny when examined in the light of modern economic theory.

I will not take a stand on whether or not cluster policy is 'right' or 'wrong', nor will I try to convince readers either way. We currently just do not know enough. Furthermore, there are no simple and universal answers to these questions, either on the academic side or on the policy side. However, I want to reject the idea that nothing interesting can be learned about geographic concentration because 'all clusters are special cases'. Instead, I want to push the idea that *valuable lessons can be learned*, and that it is time to think more systematically about clusters to narrow the gap between the scientific evidence that we have and the policy discourse we hear. I view this paper as one step in that direction. I want to suggest a number of avenues to explore building on new developments over the last ten years or so. My aim is to: (i) step back and provide a simple integrative overview of what we know about clustering of economic

---

[7] The following excerpt from a recent 'cluster cookbook' epitomizes this situation:

> *"This guide is written for you, as a policy maker wondering whether a cluster policy could be beneficial to your territory and if so what it should be like in order to achieve your strategic objectives. This handbook will not fully answer your questions, but it does provide commented experiences of good and less good practice of decision makers in comparable situations to yours. It will thus allow you to answer your questions yourself. Apart from policymakers, cluster managers, cluster researchers and other people interested in clusters will find food for thought in this guidebook."* ("Clusters and clustering policy: A guide for regional and local policy makers", European Union, 2010, p.3).

activity; (ii) address how we can think about the main ingredients in a *systematic* way; and (iii) highlight how we can put to use a number of concepts that play a key role in delineating clusters, understanding some of their causal mechanisms, and assessing their causal effects on the outcomes that have captured so much of the policy debate.

I will approach these questions in a quantitative way through the lense of 'coagglomeration patterns', i.e., the tendency of industries to locate together. Those patterns are valuable tools that allow one to look at most aspects of clusters through a common lense. I will explain how different economic mechanisms lead to productivity gains for certain industry pairs when they are geographically close. For example, industry pairs that are strongly linked by vertical buyer-supplier relationships, and which face high transport costs for their intermediates, will tend to locate together to reduce those costs. As a first step, we hence need to understand the microeconomic mechanisms that link industries. We then have to investigate whether those mechanisms are indeed a *cause* for geographic clustering. This is a complicated task since many confounding factors — the uneven distribution of natural resources, the proximity to final demand, the presence of other industries, of infrastructure, or past factors that were once relevant but are no longer today — can lead industries to be geographically clustered. Finally, I will suggest how we can think about the causal effects of clustering on various economic outcomes. The identification challenges at small geographic and industrial scales are severe, but coagglomeration patterns can help one to make progress there.

The remainder of this paper is structured as follows. In Section 2, I review what we know about the coagglomeration of industries. I first explain the underlying theory in a diagrammatic way, give some details on measurement issues, and provide some evidence for Canada. In Section 3, I show how coagglomeration patterns help us to delineate clusters, investigate their underlying causal mechanisms, and are potentially useful to construct instruments that allow us to look at the causal effect of clusters on economic outcomes. I finally discuss some open questions and current research frontiers in Section 4.

2. Coagglomeration: Theory, measurement, and evidence

The central element of this paper is the *coagglomeration patterns* of industries. Coagglomeration — a term coined by Ellison and Glaeser (1997) in their seminal paper on measuring industrial localization, and used for the first time empirically in Dumais, Ellison, and Glaeser (1997, 2002) — refers to the tendency of industries to locate together. I will make clear later how we can measure coagglomeration, how the *relatedness* of industries should matter, and how it can be used in the cluster context to: (i) delineate clusters; (ii) better understand the causal agglomeration mechanisms underlying them; and (iii) construct instruments that allow us to better understand the effects of clusters on outcomes. Before doing this I will briefly review the theory, discuss measurement issues, and provide evidence on coagglomeration in Canada.

2.1 *Theory*

The analysis of coagglomeration is a recent development in urban economics and economic geography. Helsley and Strange (2014) point out that the term coagglomeration was completely absent from the literature in the mid-2000s: it does not appear in the *Handbook of Regional and Urban Economics, vol.4*, or in Fujita and Thisse (2002), the two authoritative sources on agglomeration during that period. Since then, a few empirical studies have harnessed the power of coagglomeration to investigate agglomeration patterns and their determinants.[8] The economic theory of coagglomeration is even more sparsely developed than the empirics (see Ellison and Glaeser, 1997; Helsley and Strange, 2014; Kerr and Kominers, 2015). The reason is that most of the literature on agglomeration has either focused on complete specialization of places — due to *localization economies* that operate only within industries — or on complete diversification — due to *urbanization economies* that operate symmetrically across all industries. This state-of-affairs reflects the classical 'Marshall-vs-Jacobs' debate that has underpinned the question on the sources of agglomeration economies — and their relative strength — since Henderson's (1974) seminal contribution. The reality is of course far more complex since "no city is really a one-industry town, not even Hollywood or the Silicon Valley. Neither is any city simply a share of the diverse national population. New York's diversity is different from that of Los Angeles." (Helsley and Strange, 2014, p.1064). Though obvious, this insight has been neglected, mainly because it is hard to model. Analyzing non-trivial location patterns where multiple industries can sort across multiple locations is an inherently difficult problem (Behrens and Robert-Nicoud, 2015; Gaubert, 2016). The problem is made difficult by the fact that both the sizes and the composition of locations are endogenous, whereas productivity depends on both of these variables. This makes the problem hard to tackle under a general structure of possibly asymmetric externalities across multiple industries. Also, the presence of externalities and non-convexities allow for many possible equilibria.

To expose the fundamentals of the theory of coagglomeration, I take the simplest possible approach in the line of Henderson (1974) and Helsley and Strange (2014). Following Duranton (2011), I provide a simple diagrammatic exposition. Assume there are $\mathcal{I}$ perfectly competitive industries that can potentially operate in $\mathcal{C}$ clusters. For simplicity, suppose that labor is the only factor of production. It is perfectly mobile across clusters but specific to industries. I can relax that point but let me choose the simplest possible case. Let $L_c = \sum_i L_{ic}$ denote the size of cluster c, where i indexes industries. Output of industry i in cluster c is $Y_{ic} = A_i(\mathbf{T}_c)L_{ic}$, where L_{ic} is the labor used by that industry, and $\mathbf{T}_c$ captures 'cluster scale and composition'. The latter, which is potentially a vector, shifts production via the industry-specific function $A_i(\cdot)$. I refer to $A_i(\mathbf{T}_c)$ as the 'agglomeration externalities', which generate external increasing

---

[8] See, among others, Duranton and Overman (2005, 2008); Ellison, Glaeser, and Kerr (2010); Faggio, Silva, and Strange (2014); Helsley and Strange (2014); Howard, Newman, and Tarp (2015); Delgado, Porter, and Stern (2016a,b); O'Sullivan and Strange (2016); and Behrens and Guillain (2016).

returns for firms. These externalities arise due to various sharing, matching, and learning mechanisms that operate between geographically proximate firms. An example of *sharing externalities* is provided by either infrastructure investments (e.g., a container terminal), or a wider range of specialized local intermediates produced under increasing returns. Both can only be supplied locally if there is a sufficiently large demand that allows firms to cover their fixed costs. An example of *matching externalities* is given by the thicker labor markets that agglomerations can sustain, and which improve both firm-worker matches and provide insurance against adverse idiosyncratic shocks. Last, knowledge spillovers that can occur either accidentally between firms or by labor mobility and labor poaching in a cluster, are an example of *learning externalities*. See Duranton and Puga (2004), Puga (2010), and Behrens and Robert-Nicoud (2015) for more references and detailed reviews of the mechanisms underlying these externalities.[9]

Assuming that firms and labor markets are competitive, the wage in industry i and cluster c is given by $w_{ic} = y_{ic} = A_i(\mathbf{T}_c)$, which depends on the agglomeration externalities. Generally, productivity and wages increase with some measure of cluster size. If this was the end of the story, clusters would always grow. Yet, living in a cluster of size L_c entails costs — land rents, congestion, pollution — of $C(L_c)$, which I assume are an increasing function of cluster size only (but not of its composition).[10] Note that all standard urban models generate an increasing and convex urban cost function (Fujita, 1989; Duranton and Puga, 2015). I refer to these costs as cost-of-living for short. Observe that in more complex settings, $C(L_c)$ may be a non-monotone function of the local population. For example, $C(L_c)$ could be a reduced-form that subsumes both urban costs and different urban amenities. The latter may be partly endogenous to cluster size. Some amenities — movie theaters or opera houses — are only provided when the local population is large enough. In that case, the cost curve C may have a more complex and potentially non-monotonic shape: when people leave small places, they may reduce their attractiveness if the movie theater shuts down, though real estate prices and congestion may decrease. In that case, some places may completely disappear and the model may have multiple equilibria, as in the 'new economic geography' literature (Fujita, Krugman, and Venables, 1999), since migrants do not internalize the depressing effect they have on the communities they leave.

---

[9] Signalling may be another important mechanism. Firms may choose to locate in a cluster because it is 'the place to be', thereby signalling their innovativeness or productivity.

[10] Generally, the costs of clustering may be industry-specific, e.g., if wages across industries are not equalized or if industries use different amounts of land. Returning to the example of matching, only industries and workers for which matching is more important — or industries that face greater uncertainty and 'competitive instability' (Strange, Hejazi, and Tang, 2006) — will pay these higher costs and tend to coagglomerate. See Davis and Dingel (2014) and Gaubert (2016) for recent treatments of this problem using the tools of monotone comparative statics, and Behrens and Robert-Nicoud (2015) for a survey.

Mobile agents are attracted to places that offer them the highest net benefits. Let

$$V_i(\mathbf{T}_c, L_c) \equiv w_{ic} - C(L_c) = A_i(\mathbf{T}_c) - C(L_c) \tag{1}$$

be the net benefit of locating in cluster c for an agent employed in industry i. I denote by $S(V_i)$ the labor supply curve of industry-i workers, which corresponds to the number of those workers drawn to the cluster when it offers net benefits V_i. While some agents are very mobile across locations, others are completely immobile. More generally, it is reasonable to assume that as clusters expand, the labor supply becomes more inelastic: the most mobile agents move first to exploit small wage differentials, whereas subsequent growth requires larger wage increases to overcome the moving cost of the least mobile agents (see, e.g., Murata, 2003, for a formal model). A location equilibrium is such that the supply of workers matches the demand for workers in each industry in each cluster — conditional on the net benefits in the cluster — and that no worker can profitably switch location.

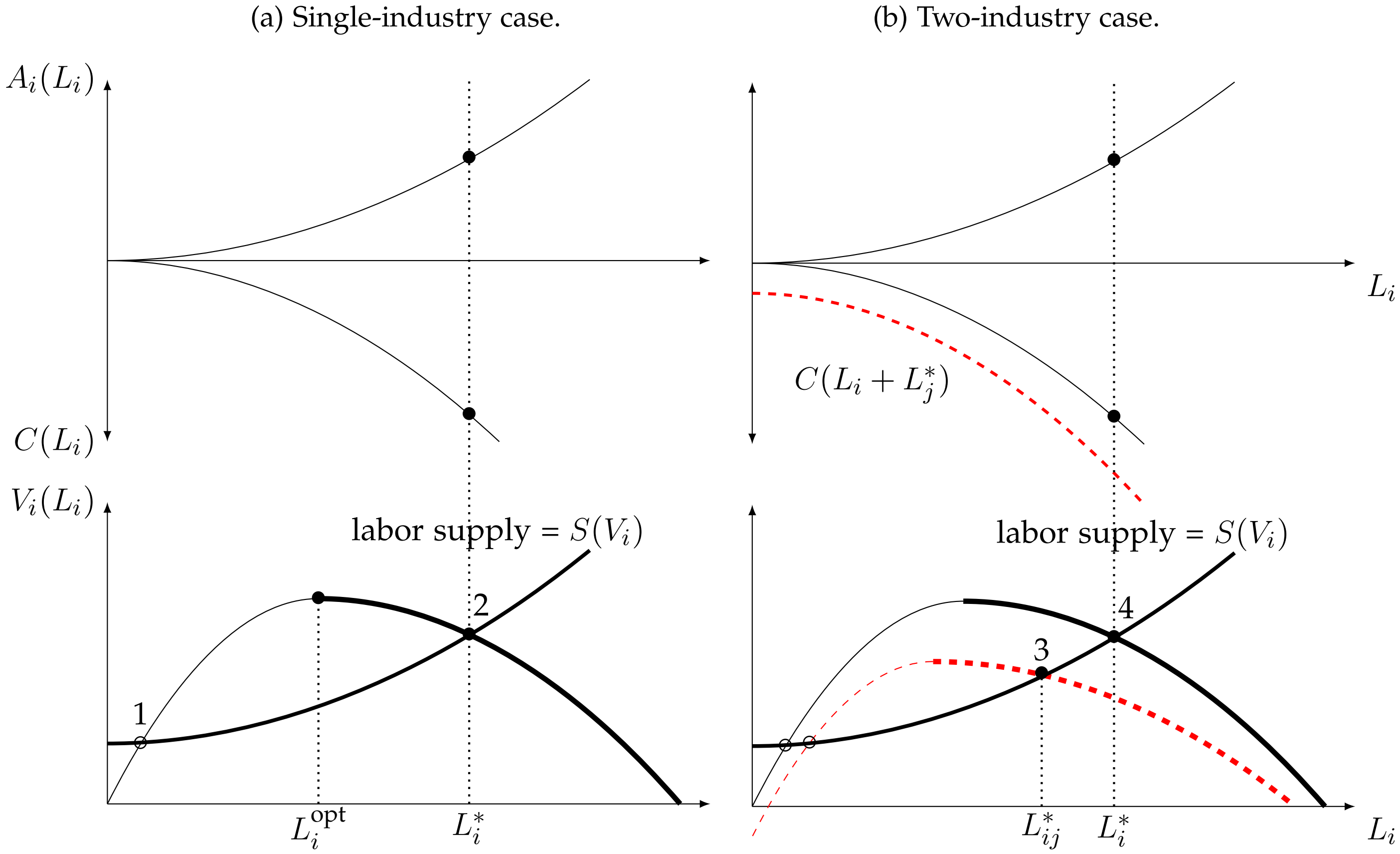
What is the relevant variable $\mathbf{T}_c$ in (1)? This is an empirical question, and it is far from settled. In most of the literature, $\mathbf{T}_c$ has been taken as a scalar, usually $\mathbf{T}_c = L_{ic}$. This corresponds to the case of localization economies and leads clusters to specialize. Conversely, a part of the literature has let $\mathbf{T}_c = \sum_i L_{ic} = L_c$, which corresponds to the case of symmetric urbanization economies that lead clusters to perfectly diversify. In fact, most cities are neither one-industry towns nor scaled replicas of the national industrial structure. This means that we need to move beyond these two extreme cases: cities and clusters are *specialized in different mixes of industries*. Helsley and Strange (2014) thus consider that $\mathbf{T}_c = (L_{1c}, L_{2c}, \ldots, L_{Ic})$ is a function of the exact composition of local employment. This captures the idea of 'co-agglomeration of related industries' and allows one to examine the cases in which clusters are specialized in different industrial mixes. In what follows, I refer to these economies as *coagglomeration economies*, to contrast them with both localization and urbanization economies. Observe that this specification captures both localization and coagglomeration (an asymmetric form of urbanization) economies by $\partial A_i / \partial L_{ic} > \partial A_i / \partial L_{jc} > 0$: productivity is increasing in own-industry employment, as well as in other-industry employment. The inequality is assumed to hold since localization economies seem to be empirically stronger than urbanization or coagglomeration economies, though more work is needed here.[11] Note that the above inequality implies, almost by definition, that clusters will generally be asymmetric, in contrast to the case where $\mathbf{T}_c = \sum_i L_{ic}$ which generates symmetric clusters in terms of their composition.

Panel (a) of Figure 2 illustrates the baseline case with a cluster that specializes in a single industry i because of localization economies. The upper part of panel (a) plots the wage curve, $A_i(L_i)$, in the positive part, and the cost-of-living curve, $C(L_i)$, in the negative part. Because

---

[11] Most of the empirical evidence suggests that own-industry spillovers (localization economies) are stronger than (symmetric) cross-industry spillovers (urbanization economies). See Henderson (2003); Henderson, Kuncoro, and Turner (1995); Rosenthal and Strange (2004); and Duranton et al. (2010, p.101). I am not aware of any work looking specifically at 'coagglomeration economies'.

Figure 2: Baseline cases with one or two industries and no coagglomeration economies.


(a) Single-industry case.

(b) Two-industry case.

we look at a single cluster, we omit the cluster index c and equate cluster size L_c with industry size L_i to alleviate notation. The net benefits curve, $V_i(L_i)$, is depicted in the lower part of panel (a), and it has the standard ∩-shaped form. An equilibrium occurs where the net benefits curve intersects the labor supply curve: at that point, the net benefits associated with cluster size L_i attract exactly L_i workers. Note that point 1 (hollow circle) is not a stable equilibrium, since the net benefits the cluster offers increases faster than those required to attract more workers. Stable equilibria necessarily occur to the right of the optimal cluster size L_i^{opt}, i.e., clusters are too large at a stable equilibrium (Henderson, 1974). Point 2 (solid circle) is a stable equilibrium.

Panel (b) of Figure 2 illustrates the case of two industries, i and j, locating in the same cluster but without coagglomeration economies. I vary L_i holding L_j fixed at L_j^* to simplify the presentation. Putting together the two different industries in the same place increases the cost-of-living, $C(L_i + L_j^*) > C(L_i)$, but yields no additional benefits. Consequently, the new net benefits curve, depicted by the red dashed line in the bottom part of panel (b), shifts downwards, and the employment of industry i conditional on L_j^* (denoted $L_{ij}^* \equiv L_i(L_j^*)$ at point 3), shrinks.

Some comments are in order. First, we can see that the coagglomeration of i and j is not efficient. In the case where only industry i would operate in the cluster, the net benefits would be higher and industry employment in i larger (point 4 associated with L_i^*, and not point 3 associated with L_{ij^*}). A symmetric argument applies for industry j. Hence, both industries would be better off in clusters specialized solely in their own type. Why can coagglomeration still be an equilibrium? Because of the *first-mover disadvantage:* no firm in industry i can deviate to start a new cluster specialized in industry i only, since the net benefits it could offer at $L_i = 0$ would be strictly smaller than that in the existing cluster. Hence, some 'large scale' coordination is required to move industry i to a new location that 'jump starts' local operations at a sufficiently large scale. Decentralized individual migration decisions cannot solve this coordination failure. Thus, inefficient coagglomeration may be sustained in panel (b) of Figure 2 because of the own size of industry i in the cluster — which generates enough agglomeration benefits compared to locations with less of the own industry — plus the first-mover disadvantage that requires that there be coordination in relocation (this is sometimes called the 'migration pathology'; see Helsley and Strange, 2014).[12] Of course, since the case I just discussed is clearly inefficient, the question arises as to why we should see it in the first place?

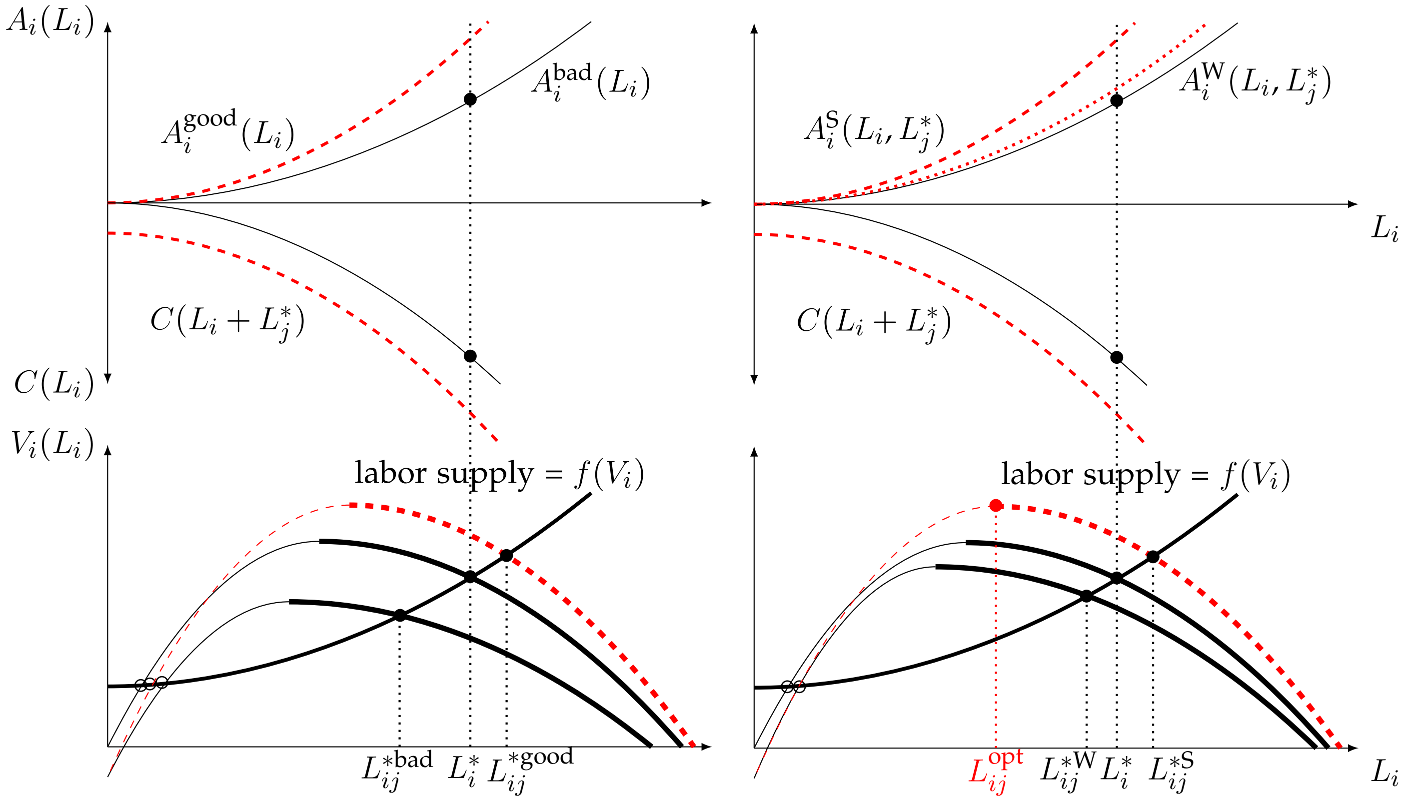
Panel (a) of Figure 3 depicts one example of why we could see inefficient coagglomeration. Assume that there are two sites: a 'bad' site, which yields agglomeration benefits A_i^{bad} for industry i; and a 'good' site, which yields agglomeration benefits A_i^{good}.[13] I continue to assume, for the moment, that there are no coagglomeration economies, i.e., neither of the two A_i's depends on the presence of another industry. The red dashed benefits curve lies above the black in the upper part of panel (a): for any given size L_i, conditional on L_j^*, the benefits are higher on the good site than on the bad site. As depicted in panel (a) of Figure 3, the coagglomeration of i and j at the good site may occur just because it yields higher benefits to industry i (upper red dashed curve); and inefficient coagglomeration at the good site may actually be better than specialization at the bad site (middle black curve). But we could equally well see coagglomeration at the bad site, which could also be sustained (lower black curve). The latter case might seem weird, but it could be rationalized in a more dynamic context because site quality (and the A_i's) are not fixed through time. Coagglomeration at sites with a natural advantage may be initially efficient, but that advantage can disappear at some moment with the industries remaining coagglomerated, even though it would now be more efficient to have them operating at separate sites (see, e.g., Bleakly and Lin, 2012, for evidence on the historical lock-in of economic activity to sites). New York's natural harbor probably played an important

---

[12] One aspect that I disregard in the simple model is the sunk nature of some capital investments like buildings and infrastructure, which tend to make agents less mobile. With non-malleable investments, some firms are sunk on location and we can get lock-in of agglomeration patterns. In that case, it is not enough for firms to be individually small and to allow for coordinating devices to potentially improve on the market outcome (see Henderson and Venables, 2014).

[13] Surprisingly, urban systems models with site heterogeneity have not been extensively studied until now (see Albouy, Behrens, Robert-Nicoud, and Seegert, 2016 for an exception).

Figure 3: Cases with locational advantage or coagglomeration economies.



(a) Locational advantage.

(b) Coagglomeration benefits.

role for many industries in the past, and they coagglomerated there to share that locational advantage. These patterns can then persist for a long time, though it would be more efficient to not have all those industries remaining in New York. Yet, they may be stuck. Note that the foregoing argument can easily be extended to the case of cluster policies. Assume that policy makers think that there are potential complementarities across industries (although there are none, or they are very small). It may be tempting to implement a place-based cluster policy that targets both i and j. This is akin in its effect to natural advantage. The same fundamental questions and problems arise. Once the industries are clustered — due to the incentives given by the cluster policy — if those incentives disappear, public policy may have locked industries into suboptimal configurations and places. This is dangerous and potentially impossible to undo. Note finally that even if coagglomeration can be efficient in my example, because of site heterogeneity, it is *spurious* in the sense that the coagglomeration itself is irrelevant: it is driven by locational advantage, and efficiency considerations of individual industries dictate that the industries end up in the same place, but it is not because industries end up in the same place that they become more efficient. This already hints at identification issues that need to be carefully considered in empirical work.

Panel (b) of Figure 3 depicts the case with coagglomeration economies. Assume that those economies can be 'strong' (dashed red line, superscript 'S'), or 'weak' (dotted red line, superscript 'W'). The lower part of panel (b) shows that the efficiency of coagglomeration depends on whether coagglomeration economies are strong or weak. In the weak case, we may (again) end up with inefficient coagglomeration. Note that this is obvious, since I have argued before that it may even occur in the complete absence of coagglomeration economies. Hence, even if coagglomeration may give rise to productivity gains, it may be inefficient if the urban costs it generates more than outweigh the coagglomeration benefits. Taking into account those costs is often neglected in analyses dealing with clusters (see Duranton et al., 2010, and Duranton, 2011, for discussions).

To summarize the key insights, the analysis of coagglomeration "[...] sharply challenges the conventional wisdom that the size and composition of cities are necessarily driven primarily by agglomerative efficiencies [...] equilibrium cities will be inefficient in scale and may be inefficient in composition as well." Helsely and Strange (2014, p.1065). While the inefficiency in scale due to the 'migration pathology' is well known since at least Henderson (1974), the inefficiency in cluster composition is new — though somehow expected. Beneficial coagglomeration need not occur, and inefficient coagglomeration, that does not reflect any coagglomeration economies, may occur.[14] Specialization may persist even if coagglomeration is beneficial since specialized places offer advantages based on size, and no industry/firm wants to relocate first because it will lose localization economies linked to own-industry size. Conversely, the coagglomeration of unrelated industries may occur if: (i) there are locational advantages common to two unrelated industries (e.g., coastal locations for 'Shipbuilding' and 'Seafood Preparation and Packaging'); and (ii) based on the 'migration pathology' and the first-mover disadvantage. If, for whatever reason, there is enough employment of industry i and industry j in cluster c, the importance of own-industry size may lead to the persistence of colocation of the two industries, not because they benefit from each other but because no one is willing to be the first to relocate to some smaller place, thus forfeiting the benefit of own

---

[14] A note on the Pareto efficiency of coagglomeration: In the single-industry case, we can have efficient cluster sizes (e.g., if developpers in the spirit of Henderson, 1974, restrict cluster sizes to L_i^{opt}; see panel (a) of Figure 2). This in not possible with coagglomeration. Consider $L_{ij}^{*\text{S}}$ in Figure 3 first. This cannot be Pareto optimal since both industries are too large in the cluster. Reducing the size of industry i will improve net benefits of industry-i agents, and since $\partial A_i/\partial L_i > \partial A_j/\partial L_i$, while urban costs change in the same way, this must also be beneficial to industry-j agents. Thus, clusters are too big at a stable migration equilibrium and restricting cluster sizes can improve outcomes for everyone. Assume next that the cluster is industry-i optimal (i.e., we are at L_{ij}^{opt} for industry i). Because it is industry-i optimal, any marginal change in L_i^* has a second-order effect on industry-i agents, whereas it has a first-order effect on industry-j agents. As Helsley and Strange (2014) show, *when* $\partial A_i/\partial L_{ij}^* > \partial A_j/\partial L_{ij}^*$*, no stable equilibrium can be Pareto efficient*. The reason is that workers always favor more of their own type and want less of the other type. The assumption that $\partial A_i/\partial L_{ij}^* > \partial A_j/\partial L_{ij}^*$, which implies $\partial V_i/\partial L_{ij}^* = \partial A_i/\partial L_{ij}^* - C'(L) > \partial A_j/\partial L_{ij}^* - C'(L) = \partial V_j/\partial L_{ij}^*$ is fundamental, otherwise decreasing L_{ij}^* (which increases V_i by reducing urban costs more than agglomeration benefits) may be detrimental to industry-j agents. In that case, the coagglomeration of economic activity may lead to Pareto efficient equilibria. Yet, as Helsley and Strange (2014) argue, this configuration is unlikely to be relevant given the empirical evidence at hand (see footnote 9).

size. Historic patterns then create path dependence and lock economic activity into inefficient coagglomeration patterns.

In the end, coagglomeration economies allow for a wide range of possible equilibrium configurations. This raises the question of whether or not coagglomeration is a useful tool for empirical work, both to learn something about the causal mechanisms, and to look at the effects of clusters on (beneficial) outcomes. I will argue below that — despite all these issues that we need to keep in mind — the answer is 'yes'. Coagglomeration patterns are useful, and the theory helps us to both understand which crucial factors we need to control for and invites us to be careful in our analysis.[15]

2.2 *Measurement*

The literature has identified several ways to measure coagglomeration. Some of the measures are intuitive but lack theoretical foundations (e.g., location quotients or simple covariances of employment across locations; see, e.g., Porter, 2000, and Delgado, Porter, and Stern, 2016a,b). More recently, better measures with theoretical foundations have been devised. The most well-known is the Ellison-Glaeser (EG) index of coagglomeration (Ellison and Glaeser, 1997), which looks at employment correlations across industries in the same locations, controlling for the size distribution of employment across plants in the industry. One limitation of that measure is that it has to be computed using predefined spatial units. Duranton and Overman (2005, 2008) propose another measure — the Duranton-Overman (DO) index — to overcome that limitation. Their measure is based on the kernel-smoothed distribution of bilateral distances across plants or employees in industries. I will use this point-pattern based measure since I can draw on detailed microgeographic (geocoded) data that obviates the need for predefined spatial units for the analysis.[16] Note that all empirical measures that we have are, at best, only loosely related to the theory of coagglomeration. They are derived from statistical processes and largely lack microeconomic foundations (see, e.g., Ellison and Glaeser, 1997, for the EG index; and Kerr and Kominers, 2015, for the DO index).

Figure 4 provides a simple example. As panel (a) shows, the distribution of plants in 'Motor Vehicle Manufacturing' (NAICS 3361) and in 'Motor Vehicle Parts Manufacturing' (NAICS 3363) seem to be visually close (the blue and red dots are colocated). Panel (b) — which depicts the cumulative distribution of bilateral distances between the plants in the two industries, weighted by the plants' employment — shows that these two industries are indeed on average

---

[15] Recent simulation work by O'Sullivan and Strange (2016) suggests that beneficial coagglomeration tends to be reflected in the equilibria of models of coagglomeration. According to their simulations, coagglomeration patterns do reflect — on average — coagglomeration economies. I will come back to this point below and show that coagglomeration patterns indeed do embody (at least partly) mechanisms that seem to be important for describing the benefits of coagglomeration. This is an important result for empirical work.

[16] Howard et al. (2015) have recently proposed another measure, the 'excess colocation index'. That measure is very similar in spirit to the existing measures.

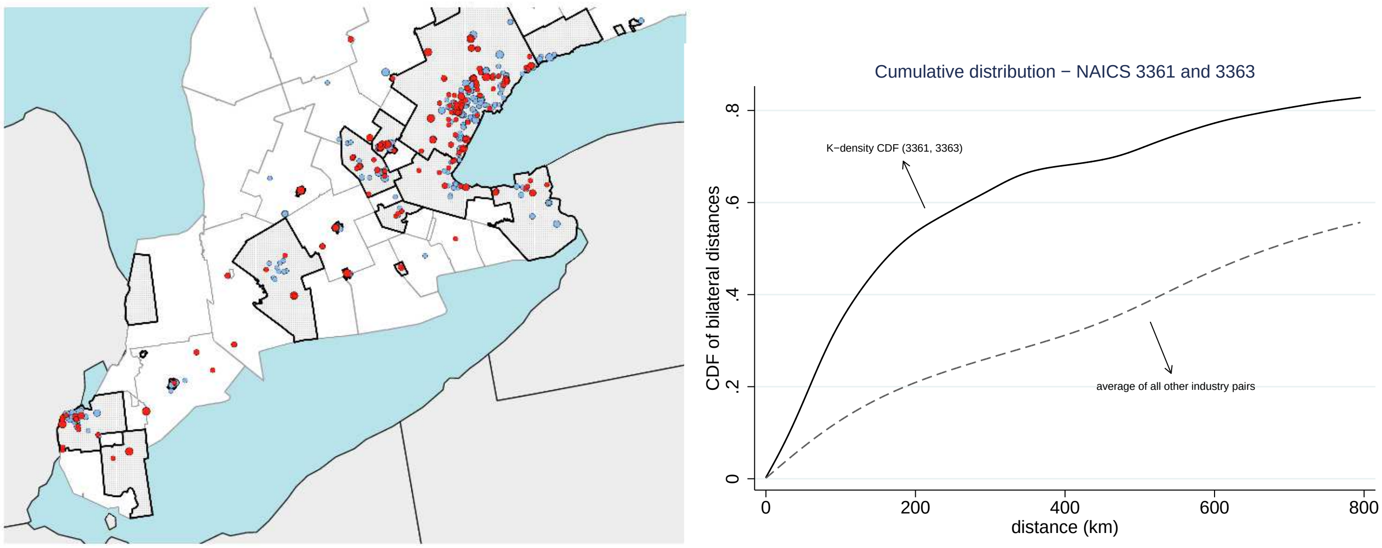


Figure 4: Coagglomeration patterns in the automotive industry.

(a) Raw geographic distribution.

(b) CDF of bilateral distances.

much closer geographically than plants from any other random pair of industries (the dashed grey line in the right panel). More precisely, about 9.06% of bilateral distances are less than 25 kilometers for the pair (3361, 3363), whereas the average share of distances below 25 kilometers for all other pairs of industries is only 3.92%. Thus, there appears to be substantial coagglomeration for these two automotive industries.[17]

In what follows, I will use the CDF of bilateral distances between plants — weighted by employment size — in two industries at some given distance threshold (e.g., 25 kilometers) as a measure of the extent of coagglomeration between those industries. I denote by c_{ij}^t that measure for industries i and j at time t. One simple interpretation is as follows. Take the foregoing example, where 9.06% of bilateral distances are less than 25 kilometers for the two automotive industries, but only 3.92% are below that threshold on average for all other industry pairs. To a first-order approximation this means that if I draw at random a pair of employees in the industries (3361, 3363), and at random a pair of employees in any other two industries, then the probability that the former are less than 25 kilometers apart is about 9%, whereas it is less than half as much in the second case.

---

[17] Statistical tests based on random permutations can be used to formally test the significance of the coagglomeration patterns (see Duranton and Overman, 2005, 2008). These tests use the overall distribution of the two industries as a benchmark and randomly reshuffle the industries' plants across sites occupied by the industries. Then, the procedure tests whether the observed coagglomeration patterns significantly depart from the confidence regions generated using 1,000 random permutations under the benchmark.

2.3 *Evidence*

Evidence about the extents and magnitudes of coagglomeration have been established for a few countries only. Ellison and Glaeser (1997), and Ellison et al. (2010) show that coagglomeration is fairly widespread in U.S. manufacturing industries. Duranton and Overman (2008), and Faggio et al. (2015) report similar finding for the U.K. And Howard et al. (2015) identify patterns of coagglomeration in Vietnam. I now apply the Duranton-Overman microgeographic measure described in Section 2.2 to Canada and document a number of summary figures on the extent and magnitude of agglomeration and coagglomeration for manufacturing industries.[18]

How widespread is the geographic concentration of manufacturing industries in Canada? Table 1 summarizes the agglomeration patterns of 4-digit manufacturing industries and the coagglomeration patterns of 4-digit manufacturing industry pairs in Canada in 2001 and 2007, respectively.[19]

Table 1: Shares of significantly agglomerated industries and coagglomerated industry pairs.

	Agglomeration: 4-digit mfg industries		
Year	% agglomerated	% dispersed	% random
2001	**56.47**	9.41	34.12
2007	**50.59**	12.94	36.47
	Coagglomeration: 4-digit mfg industry pairs		
	% agglomerated	% dispersed	% random
2001	**60.67**	9.55	29.78
2007	**46.72**	13.53	39.75

Notes: Results for 85 4-digit NAICS industries, and $(85 \times 84)/2$ unique 4-digit industry pairs. Significance tests are based on counterfactual distributions using 200 random permutations (see Duranton and Overman, 2005).

As Table 1 shows, about 60%–70% of industries display a non-random pattern in the data, i.e., they are either significantly agglomerated or significantly dispersed. Coagglomeration seems to be pervasive and widespread in Canada, with about 50%–60% of industry pairs significantly coagglomerated.[20]

Figure 5 depicts the number of significantly agglomerated (panel (a)) and significantly coagglomerated (panel (b)) industries and industry pairs by distance in Canada in 2007. The figure shows that the agglomeration and coagglomeration patterns differ in important ways. Own-industry agglomeration tends to occur predominantly at short distances (note

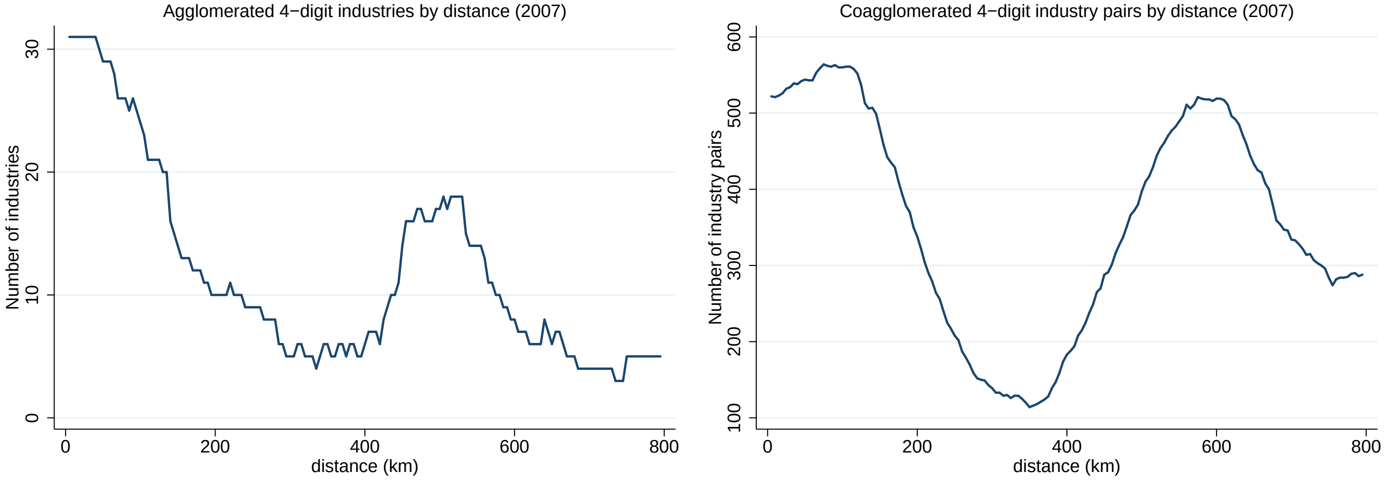
---

[18] For reasons of data availability, I only discuss the case of manufacturing industries. Evidence of coagglomeration for services is very scarce to date (see Kolko, 2010, for direct evidence on services; and Gabe and Abel, 2016, for the coagglomeration of occupations across manufacturing and service industries).

[19] There is less coagglomeration at the 6-digit NAICS level. The same holds true for agglomeration patterns of individual industries (see Behrens and Bougna, 2015, and Behrens and Guillain, 2016, for details).

[20] The extent and significance of coagglomeration seems to be decreasing over time. This concurs with other studies that have documented that own-industry localization is progressively falling in Canadian manufacturing (see Behrens and Bougna, 2015; and Behrens, Bougna, and Brown, 2015).

Figure 5: Spatial agglomeration and coagglomeration profiles in Canadian manufacturing.


(a) Agglomeration by distance.

(b) Coagglomeration by distance.

the large first peak in panel (a)). In contrast, coagglomeration patterns involve a larger number of industry pairs that are at longer distances (note the large second peak in panel (b)). The first coagglomeration peak occurs at around 100 kilometers, which suggests that many industry pairs are significantly coagglomerated at larger geographic scales. This would be consistent with a pattern where buyer-supplier relationships that operate at larger distances are important. The smaller second peak for agglomeration — and the larger second peak for coagglomeration — are consistent with a pattern in which different cities specialize in different industries: the industrial mix of Toronto is different from that of Montréal, thereby leading to many industry pairs that are significantly overrepresented at a distance around 600 kilometers. Yet, there are still a substantial number of industry pairs that locate in close geographic proximity, as can be seen from the first peak in panel (b), and as clearly depicted in panel (a) of Figure 4. In a nutshell, coagglomeration patterns are heterogenous and likely to contain various pieces of important information. Note that Figure 5 is purely descriptive and carries no information on what is driving the observed patterns. The second peak, in particular, is unlikely to contain information that is useful to identify the causal mechanisms driving agglomeration and coagglomeration. Those mechanisms operate at shorter distances, so that my focus will be on short-distance patterns in what follows.

Table 2 shows the top-ten agglomerated industries in Canada in 2005, as well as the top-ten industries that are coagglomerated with 'Motor Vehicle Manufacturing' (NAICS 3361), using a distance threshold of 25 kilometers. As can be seen, 'Motor Vehicle Manufacturing' is among the most strongly geographically concentrated industries in Canada, with about 16% of bilateral distances between plants below 25 kilometers. Other industries that display strong geographic concentration patterns are, e.g., related to textiles (see panel (b) of Figure 1). As Table 2 shows, 'Motor Vehicle Parts Manufacturing' (NAICS 3363) is the most strongly coagglomerated industry with 'Motor Vehicle Manufacturing' in Canada in 2005. This does

Table 2: Agglomeration and coagglomeration of 4-digit NAICS industries in 2005.

NAICS	CDF, 25km	Industry name	CDF, 25km	Industry name
3335	28.33%	Metalworking Machinery Manufacturing		
3321	22.40%	Forging and Stamping		
3152	17.77%	Cut and Sew Clothing Manufacturing		
3361	15.62%	**Motor Vehicle Manufacturing**		
		Top ten industries that 'Motor Vehicle Manufacturing' is the most coagglomerated with —	18.32%	**Motor Vehicle Parts Manufacturing**
			12.10%	Cement and Concrete Product Manufacturing
			10.46%	Fabric Mills
			10.39%	Footwear Manufacturing
			10.13%	Textile and Fabric Finishing and Fabric Coating
			9.91%	Other Fabricated Metal Product Manufacturing
			9.30%	Plastic Product Manufacturing
			7.31%	Textile Furnishings Mills
			6.80%	Other Miscellaneous Manufacturing
			6.61%	Machine Shops, Turned Product, and Screw, Nut and Bolt
3372	11.60%	Office Furniture (including Fixtures) Manufacturing		
3151	10.62%	Clothing Knitting Mills		
3344	9.74%	Semiconductor and Other Electronic Component Mfg		
3222	7.07%	Converted Paper Product Manufacturing		
3132	6.56%	Fabric Mills		
3328	6.00%	Coating, Engraving, Heat Treating and Allied Activities		

Notes: Agglomeration figures are from Behrens and Bougna (2015), and coagglomeration figures are from Behrens and Guillain (2016). All figures are weighted by plants' employment sizes. The results I report are computed for a 25 kilometer distance threshold.

not come as a surprise, since those two industries are strongly linked by vertical supply chains. Thus, strong input-output relationships may explain the coagglomeration of these two industries (by reducing costs, mitigating problems of supply-chain control for just-in-time inventories etc). Yet, Table 2 also reveals that some coagglomeration patterns are much less intuitive and harder to explain. They may be spurious and due to different types of locational advantage (e.g., urbanization effects or labor force effects, which could explain why 'Cement and Concrete Product Manufacturing' and 'Motor Vehicle Manufacturing' are colocated).[21] They may also be driven by the presence of '3rd industries', a point I will return to in detail below. In a nutshell, there is room for 'weird' coagglomeration patterns, which can be due to all the mechanisms I mentioned in Section 2.1. Adequately dealing with them will be important to identify the underlying economic mechanisms driving the patterns of coagglomeration. Using coagglomeration patterns to identify agglomeration mechanisms is potentially very useful, but it presents a number of important challenges that are linked to the fact that coagglomeration patterns do not necessarily carry any meaningful information.

3. Putting coagglomeration to use

Until now, I have used coagglomeration patterns in a purely descriptive way. While this is valuable by itself — to think, e.g., about the local industrial composition of regions that

---

[21] Most coagglomeration studies find some 'weird' coagglomeration pairs. Faggio et al. (2015), for example, find among the most coagglomerated industry pairs in the U.K. the 'Publishing' and the 'Jewellery and related articles' industries, which seem to have little obvious connection (except for being 'urban'). Ellison et al. (2010) find, for example, the coagglomeration of 'Handbags' and 'Photographic equipment'. All studies — including my own in Behrens and Guillain (2016) — consistently find a lot of textile, clothing, and footwear industries to be among the most coagglomerated pairs.

depends on coagglomeration forces and which determines their exposure to various industry-specific shocks — I will now explain how coagglomeration can be used as a more analytical tool to deepen our understanding of the workings of clusters. More precisely, I now show that coagglomeration patterns are useful to: (i) help define and delineate clusters, both in terms of geographical extent and industrial-functional composition; (ii) learn about the causal mechanisms driving the geographic concentration of industries; and (iii) build instruments that potentially allow us to assess the role of clusters for various economic outcomes (such as productivity and innovation, resilience, and export participation and performance) and that allow us to tackle identification issues that are hard to address with existing instruments.

Figure 6: Conceptual framework for using coagglomeration patterns.

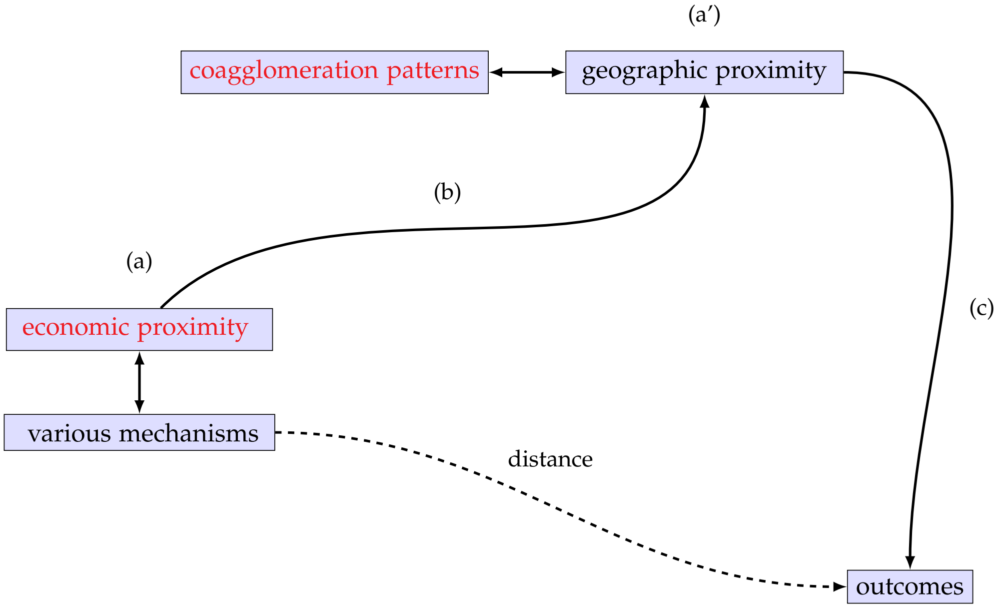


Figure 6 illustrates the three fundamental parts of my subsequent analysis. First, there are mechanisms that influence outcomes. Some of these mechanisms operate regardless of spatial proximity (think, e.g., of education and schooling or good institutions). I will be silent on these as coagglomeration cannot teach us much about them. My focus is on mechanisms that *can only operate if the economic actors are in close geographic proximity* — such us input-output sharing or knowledge spillovers. The analysis has three parts. First, in parts (a) and (a′) of Figure 6, I will briefly discuss ‘economic proximity’ and ‘geographic proximity’ and put those concepts into operation. I have briefly discussed before what the mechanisms of economic proximity could be and how they operate (see Duranton and Puga, 2004; and Behrens and Robert-Nicoud, 2015, for surveys). I addressed defining and measuring geographic proximity in Sections 2.2 and 2.3 for the specific case of coagglomeration of industry pairs. I extend this below beyond the bilateral industry pair to think more generally about clusters of related industries. Last, I will show that there is a positive association between the concepts of economic proximity that we use and the geographic proximity of industries. In other words, (a) and (a') are linked in an informative way and are to some extent the flip-side of the same coin. With these components in hand, I can then adress the remaining parts of the problem. In part (b) of Figure 6, I will explain how we can use coagglomeration patterns to analyze the *causal link* between the mechanisms of economic proximity and the geographic concentration of industries. Last, in part (c) of Figure 6, I will argue that we can also use coagglomeration patterns to better understand the causal effect of geographic concentration — or clusters more generally — on outcomes.

I will argue that we have now good theories and tools to think about part (a). We also have better tools than before to understand part (b). Yet, we have very little until now that allows us to look in an informed way at part (c), especially when it comes to empirical work. Coagglomeration patterns between industries are informative as to how we can think about points (a), (b) and (c) at the same time, using different aspects of the same tool. This makes coagglomeration patterns very valuable to the analysis of a broad range of questions related to geographic concentrations and clusters.

3.1 *Delineating clusters*

How do I know a cluster when I see one? What exactly is, for example, a 'financial', 'textile', or 'automotive' cluster? Delimiting clearly our object of analysis is vital. Though it seems intuitively clear to many observers what a cluster is, it is much harder to think about the concept of a cluster in an analytically structured way. In Porter's (2000, p.16) words, a "cluster is a geographically proximate group of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities." While this definition makes sense, it is very hard to operationalize in a *systematic* way once one starts to think seriously about it. Note my emphasis on the word 'systematic'. The objective is to get away from cluster definitions based purely on case studies, local knowledge, or 'expert judgement'. Although these components are useful — cluster studies require some qualitative components and expert assessments can be helpful sometimes — I want to focus in what follows on a purely quantitative procedure. Expert judgement unfortunately opens the door to interest-group thinking and may sometimes conflict with quantitative results that can only be better understood in a context where more disaggregated data are available.[22]

As Duranton et al. (2010) point out, three key measurement issues need to be addressed. They relate to the industrial ('group of interconnected companies and associated institutions in a particular field') and geographic ('geographically proximate') scope of clusters. They also relate to the economic mechanisms ('commonalities and complementarities'), i.e., the proximity and links between firms in non-geographic space. The first measurement problem is based on the observation that "[...] cluster boundaries rarely conform to standard industrial classifications systems, which fail to capture many important linkages across industries." (Porter, 2000, p.18).[23] It can be addressed by using a wide range of indicators of industry interactions to bundle industries into groups that are 'similar' in terms of their 'economic proximity' and, therefore, likely to interact a lot. This is now well understood and has been used in various contexts (e.g., Ellison et al., 2010; Faggio et al., 2015; Delgado et al., 2016a). The second issue is trickier, and it is directly linked to the well-know 'modifiable areal unit problem' (MAUP): "[the spatial scopes of clusters] do not conform to the boundaries of customary spatial units such as states, counties, or metropolitan areas." (Duranton et al., 2010, p.88). I will show how this problem can be addressed by using geocoded data.[24]

What follows is largely based on the cluster definition procedure of Delgado et al. (2016a), which has been used in a different variation by Behrens, Boualam, and Martin (2016).

Step 1. Grouping industries by economic proximity. The basic idea is to group industries by economic proximity (Delgado et al., 2016a).[25] The groupings can be based on various measures such as the strength of input-output links, the overlap in industries' workforce and occupational structures, patent citations between industries, labor movements across industries, or the 'internalization of activities' of one industry in another (what I call within-plant 'agglomeration', which I explain below). In what follows, I refer to the similarity measures

---

[22] Martin and Sunley (2003, p.12) propose a selection of cluster definitions. What becomes evident from the proliferation of definitions they unearth from the literature is that the 'cluster' is a 'chaotic concept'. The absence of an agreed-upon objective and easily assessable definition is highly problematic. As emphasized by Martin and Sunley (2003), this leaves the door open for almost anything. For example, consider a study that aims at assessing the benefits of clustering. If the study concludes that there are no benefits, the absence of these cluster benefits can always be rejected by cluster proponents by appealing to the idea that the clusters used have been 'misspecified' or 'mismeasured' from the start. If you do not see benefits — or so the argument goes — you have just not correctly measured the cluster. The whole idea of trying to measure cluster benefits then becomes tautological and non-falsifiable in a Popperian sense. It allows one to justify almost any type or form of cluster, depending "on what the aim of the exercise is, or the client or policy-maker for whom the analysis is intended." (Martin and Sunley, 2003, p.12). This is, at best, problematic.

[23] While this argument is not wrong, it is debatable how fundamental it is in practice. Most of the industrial 'benchmark clusters' identified in Delgado et al. (2016a) are fairly close to higher-level NAICS aggregations. See also the arguments in Duranton (2011).

[24] An unsolved problem in the literature relates to the level of industrial aggregation. To my knowledge, this point has not been seriously considered until now.

[25] Ideally, we would like to group 'plants' by economic similarity. However, the data requirements for this seem formidable and do not yet prove feasible.

between industries i and j at time t formally as $\mathbf{M}_{ij}^t$. The literature has also made extensive use of locational correlations (i.e., spatial covariances in the employment distribution of industry pairs) and sometimes coagglomeration measures c_{ij}^t. Note that the use of the latter measures — spatial covariances or coagglomeration measures — is problematic at this stage since it amounts to already *assuming* that geographic proximity reflects economic proximity. Though this is valid ex post since the data reveal that $c_{ij}^t = c_{ij}^t(\mathbf{M}_{ij}^t)$, we ideally would like to establish this result first.

The procedures used in the literature seek to group industries into 'clusters', in a mathematical sense, based on measures of economic proximity $\mathbf{M}_{ij}^t$. Industries are grouped in an optimal way into clusters of similar industries, whereas the clusters are very different from one another. Delgado et al. (2016a) use cluster algorithms to partition all U.S. 6-digit NAICS industries into 51 mutually exclusive 'benchmark cluster definitions'.[26] The groupings minimize a measure of within-group variance and maximizing a measure of between-group variance. Validation scores (likelihoods) for different cluster configurations can also be computed to find the 'best' groupings. This latter aspect is particularly interesting, since it allows for a statistical foundation for the derived clusters of related industries. Delgado et al.'s (2016a) benchmark cluster definitions include high-tech clusters such as 'Aerospace Vehicles and Defense' or 'Biopharmaceuticals', and more traditional clusters such as 'Textile Manufacturing' or 'Furniture', which look at lot like more traditional industrial classifications.

Table 3: Industry groupings for T&C, based on different 'economic proximity' metrics.

	Clusters into which the textile, apparel, footwear, and leather-related industries are grouped			
Proximity metric used	**Cluster 1**	**Cluster 2**	**Cluster 3**	Residual groupings
Within-plant complementarities	3141, **3379** Rugs and *funiture*	3131, 3149, 3133, 3132, 3159, **3231** Textile mills and *printing*	3151, 3161, 3162, 3169, 3152 Apparel and footwear	
Input-output links	**3116**, 3161, 3162, 3169 Footwear, leather, and *meat*	3131, 3132, 3133, 3141, 3149, 3151, 3152, 3159 Textiles mills, apparel and 'cut-and-sew'		
Occupational employment correlation	3131, 3132 Textile mills	3133, 3141, 3151, 3149, 3152, 3159, 3162 Textiles, apparel, and 'cut-and-sew'	3169 (a singleton cluster)	3161 (alone in one big cluster)
Knowledge flows	3161, 3162, 3169 Leather and footwear	3159, 3152 Cut-and-sew	3132 (a singleton cluster)	3131, 3133, 3149, 3141 (together in one big cluster)
Labor mobility between industries	3152, 3159, 3162 Cut-and-sew and footwear	3131, 3132, 3141, 3133, 3149, 3151, **3231** Textiles mills, rugs, hosiery and *printing*		3161, 3169 (together in one big cluster)

Notes: The four metrics that we use to cluster our 85 4-digit industries into 'economically close groups' are constructed as follows: (i) 'Within-firm complementarities' is the share of firms in a 4-digit primary industry i that report also at least a secondary code in another 4-digit industry j; (ii) 'Max input-output links' is the maximum element in the input-output tables between i and j; (iii) 'Occupational employment correlation' is the correlation coefficient between i and j's shares of 553 occupations from the US OES surveys (we exclude all occupations that report zero employment in manufacturing); (iv) 'Knowledge flows' is the use-based share of patents that originate in i and are embodied (cited) in patents of j; and (v) 'Labor mobility between industries' is the share of workers leaving industry i and moving to industry j (conditional on moving), computed using 2000-2005 CPS data that is made 'panel-consistent' using the procedure described in Madrian and Lefgren (1999). The clustering is done using the Markov cluster algorithm (MCL) by Stijn van Dongen ("A cluster algorithm for graphs. Technical Report INS-R0010, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, May 2000; see `http://micans.org/mcl/`). The graph is constructed with positive weights for all links with values above the median, and zero weights below the median, in order to allow for more variability. This is a standard procedure used to make the graph sparser to allow for meaningful groupings. See Behrens et al. (2016a,b) for additional details.

Table 3 illustrates the workings of such a cluster procedure. It identifies NAICS 3131–3169 as 'Textile and Clothing' cluster using a Markov cluster algorithm. Note that the cluster definition in Table 3 is very close to the 'Textile cluster' benchmark definition of Delgado et al. (2016a):

---

[26] Recent procedures split industries into mutually exclusive groups, based on measures of similarity. It is questionable whether cluster definitions should involve partitions of industries. Ideally, they should allow overlap to the extent that some industries contribute to the external increasing returns process in multiple industry groups.

it encompasses their four clusters 'Apparel', 'Footwear', 'Leather and related products', and 'Textile manufacturing'. The 'Leather and footwear' cluster is included in our textile cluster because, as Table 3 shows, according to the 'Within-firm agglomeration' metric, plants engaged in textile manufacturing also heavily engage in footwear and leather-related activities, whereas according to the 'Labor mobility' metric, workers are fairly mobile between the footwear and the cut-and-sew industries.[27] In any case, the clusters we identify using a different procedure and Canadian data are fairly similar to those identified by Delgado et al. (2016a) using U.S. data, which is reassuring because the two procedures seem to pick up the same relevant patterns in the data.

The fundamental final question is whether economic proximity translates into geographic proximity, i.e., whether or not $c_{ij}^t = f(\mathbf{M}_{ij}^t)$. This question is central for two reasons. First, if this is not the case then the mechanisms that we proxy using the concepts of economic proximity have no bearing on the spatial structure of the economy and are, therefore, of little interest to this type of analysis. Second, as I already stated, the 'revealed geographic proximity' of industries is often used as a criterion to group those industries in the first place (see, e.g., Porter, 1990, 2000; Delgado et al., 2016a). While this is not inherently a problem, it only makes sense if we are confident that the groupings indeed carry information about the underlying mechanisms. As Section 2.1 shows, it is far from clear in theory just how much information on coagglomeration economies is really embodied in coagglomeration patterns.

Table 4: Economic proximity is associated with geographic proximity.

Avg. share of plant pairs at less than	25km	50km
non-'textile & clothing' industry pairs	3.84%	7.37%
mixed industry pairs	3.86%	7.42%
'textile & clothing' industry pairs	5.26%	10.09%

Notes: Based on the 'textile & clothing' clusters analysis in Behrens, Boualam, and Martin (2016a,b).

Table 4 shows that pairs of plants in 'Textile and Clothing', identified as an industrial cluster of related industries in Table 3, are significantly more agglomerated than other pairs of plants. This shows that there is a positive association between economic proximity, as proxied by input-output links and other measures $\mathbf{M}_{ij}^t$, and geographic proximity c_{ij}^t.[28] Of course, that association need not be causal, as Section 2.1 shows. We return to the causal identification of

---

[27] I construct a measure of 'Within-plant agglomeration' that reports the share of plants in industry i that also report as secondary activity industry j. I return to this metric more extensively in Section 3.2.

[28] In an interesting recent contribution, Otazawa and van Ommeren (2015) connect the between-firm networks with the geographic location of firms in the Japanese city of Kure. They show that, consistent with theory, firms that are more central in buyer-seller networks (non-geographic space) are also geographically more centrally located. Atalay, Hortaçsu, Roberts, and Syverson (2011), and Bernard, Moxnes, and Saito (2015) further document that links between firms in the U.S. and in Japan are more likely to exist when they are geographically close.

the mechanisms in Section 3.2. But the results suggest that coagglomeration measures help to delineate geographic clusters in terms of industrial composition. O'Sullivan and Strange (2016) provide recent evidence using simulations of a simple coagglomeration model that — despite the large number of possible equilibria — the equilibria of the model do, on average, reflect the existence of beneficial coagglomeration of industries. Hence, measures of coagglomeration can play an important role in defining clusters, as they 'embody — on average — mechanisms'. This is potentially useful in data-poor contexts where 'economic proximity' data is either not available or of low quality (e.g., developing countries; see Howard et al., 2015, for Vietnam).

Step 2. Mapping clusters geographically. While grouping industries by economic proximity and measuring their coagglomeration patterns at the national level is informative, it reveals little about *where* clusters actually are. Once industries are grouped — and once we are confident that these groupings are indeed meaningfully associated with geographic proximity — we want to visualize clusters on the map. Various 'cluster mapping tools' exist nowadays (e.g., the U.S. Cluster Mapping Project, a joint initiative between the BEA and the Harvard Business School; see Delgado et al., 2016a,b; the Cluster mapping project of the European Union; and the Canadian 'Cluster Mapping Portal' of the Institute for Competitiveness and Prosperity). Several other countries are in the process of investing into these tools or rethinking existing tools (see, e.g., the Federal Budget 2016 in Canada, which features a 'cluster mapping project').

How can we map clusters? Two criteria are usually used to determine where to put clusters. To the best of my knowledge, there is little theoretical guidance here, so that exercise relates as much to art as to science. The two criteria are *specialization* and *size*. The former stipulates that a sufficiently large *share* of local employment is associated with the cluster. This criterion is important to rule out size effects, whereby large cities just have a larger number of employees in clustered industries. The latter criterion requires that clusters be large enough. This is important to rule out cases where a cluster consists of two firms in the clustered industry operating in a location with just three firms (in which case the location would be highly specialized in the clustered industry, though we could hardly meaningfully talk about a 'cluster'). What exactly 'sufficiently specialized' and 'sufficiently large' means is, as I said before, absolutely not clear.

Another recurring question relates to the geographic scale used for mapping clusters. Most of the existing works are constrained by the available geographic units at which the data are reported. See, for example, Delgado et al. (2016a), who work at the Economic Area level in the U.S., which corresponds to (very large) metropolitan regions. It is debatable whether this is the right unit of analysis, especially because many cluster mechanisms seem to operate quite locally, especially highly localized knowledge spillovers and other learning and human-capital externalities that attenuate very rapidly over space.[29] As noted by Martin and Sunley (2003,

---

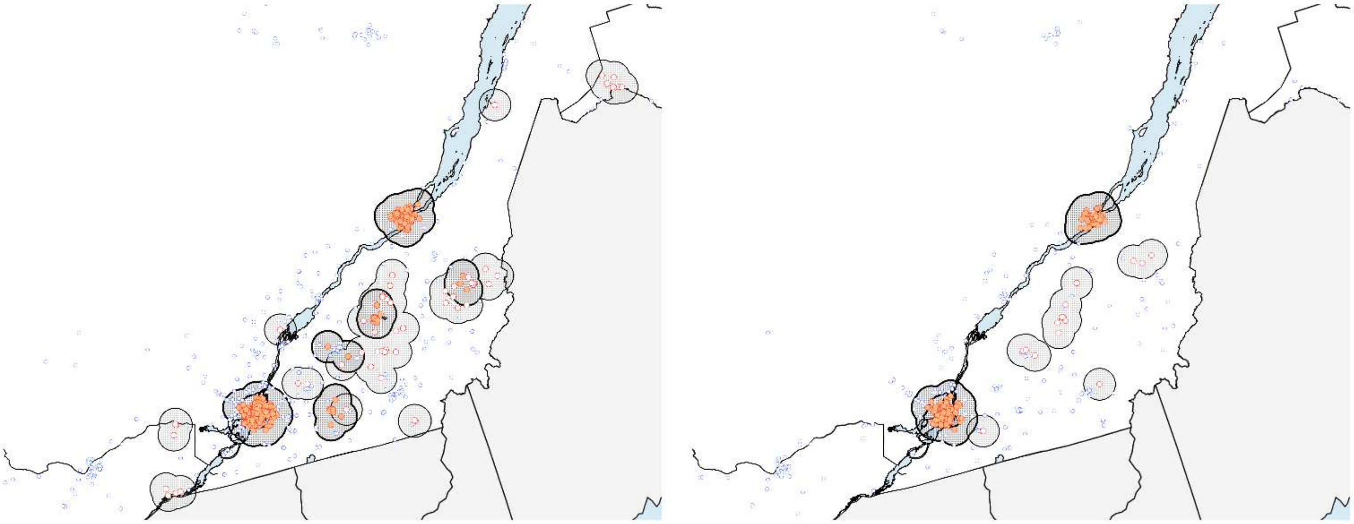
[29] Also, another number of mechanisms seem to operate at a larger international scale (Wolfe and Gertler, 2004).

p.12), "[. . .] to use the term [cluster] to refer to any spatial scale is stretching the concept to the limits of credulity." One way to get around that problem is again to use microgeographic data.

I now illustrate one way of cluster mapping for the Canadian 'Textile and Clothing' industry case, where I can work in continuous space and do not need to rely on arbitrary spatial units. Figure 7 maps the T&C clusters — based on the industrial groupings from Table 3 — for the years 2001 (panel (a)) and 2013 (panel (b)).

Figure 7: Mapping of T&C clusters in Québec and their changes over time.

(a) 2001. (b) 2013.



The technique underlying the detection of the clusters in Figure 7 is based on significance tests of the observed point patterns.[30] These techniques go back to those developed in spatial data analysis (e.g., Cressie, 1993; Diggle, 2003) to detect significant concentrations of certain characteristics (e.g., the prevalence of some disease in the population; see Duranton and Overman, 2005, and Marcon and Puech, 2010, for discussions). Loosely speaking, for each plant in the clustered industry and for some given distance threshold, we test the likelihood of observing x other plants in the clustered industry around that plant, conditional on there being y plants overall around it. We can then use draws from a hypergeometric distribution — or do some permutation bootstraps — to associate significance levels with the observed configuration around each plant. Plants that have 'abnormal' concentrations of clustered plants around them are then used as focal plants, conditional on some size criterion (at least 5 other plants).[31] We finally associate geographic buffers for the selected distance threshold around those focal

---

[30] Multi-scale core clustering, as in Buzard, Carlino, Carr, Hunt, and Smith (2015), or other approaches as in Marcon and Puech (2010) provide alternative ways to test for local clustering of economic activity. All these approaches allow for significance tests and inference as in spatial statistics to define 'hot spots' — i.e., areas of abnormal concentration — in the point patterns.

[31] An interesting exercise would be to look at entry and exit patterns, i.e., to exploit the dynamic aspects of the clustering process. See, e.g., Rosenthal and Strange (2003) for a related approach.

plants to delimit the clusters. In Figure 7, we also split clusters by size, depending on how many plants they do contain in the clustered industries (the 'core' clusters, with at least 25 plants, are highlighted in bold outlines in Figure 7).

Figure 7 illustrates two important aspects. First, as already mentioned, clusters can be highly localized and do not align well with metropolitan area boundaries or other administrative divisions. This becomes especially clear when using microgeographic data (see also Buzard et al., 2015, who delimit microgeographic clusters of R&D labs in the U.S. which do not align with any administrative boundary). Second, and less obviously, clusters change rapidly over time, even over relatively short periods (2001 and 2013 in my example). Part of this is due to profound changes in the trading environment of industries (e.g., the end of Multi-Fibre Arrangement in 2005 changed significantly the landscape of the T&C industry; see Behrens et al., 2016a,b), but part of it is also due to substantial churning of plants that does not readily show up in changes in more aggregate spatial data but can substantially change industrial patterns at a small geographic scale.

To summarize, substantial progress has been made in building better analytical tools to delineate cluster functionally and geographically. Better data on economic proximity of industries will allow us to refine the groupings of industries into coherent units. The availability of more disaggregated spatial data will allow us to better track microgeographic clusters and their evolution. Although "the statistical acrobatics employed to map 'significant' clusters are complex" (Martin and Sunley, 2003, p.21), they are no longer prohibitive. While some years ago the data were the limiting frontier, I think that theory is the greater challenge and a more pressing matter nowadays.

3.2 *Identifying mechanisms*

Identifying and understanding the mechanisms underlying the geographic concentration of industries is important for several reasons. First, it is important to assess the presence and extent of market failures. Since clusters are ripe with market failures — and since cluster policy is justified precisely by the existence of these failures — this is a first-order issue. Second, it is required to design local policies to act on market failures and various economic outcomes. Knowing whether economic activity in some industries tends to cluster because of 'knowledge spillovers' or whether this is just because of the presence of shared infrastructure is key to designing policies. Infrastructure investments might be of little help when the driving forces reflect knowledge exchanges, so the policy maker's response to address clustering in different contexts should be very different. Last, understanding the causal mechanisms is also required to frame more precisely the questions within which to identify the impact of clusters on various economic outcomes.

Coagglomeration patterns — and changes in them — are an important input for better understanding the determinants of geographic concentration. Ellison et al. (2010), Faggio et al. (2015) and Behrens and Guillain (2016) all use coagglomeration patterns to learn more about the causal drivers of the geographic concentration of industries. Coagglomeration patterns are a powerful source of evidence on the mechanisms underlying the increasing returns to scale processes that lie at the heart of the geographic clustering of economic activity.

There are many potential mechanisms and causes underlying the coagglomeration of industries: buyer-supplier relationships, workforce similarity, or knowledge sharing (Marshall, 1890); matching, sharing, and learning externalities (Duranton and Puga, 2004); or other organizational-functional (Duranton and Puga, 2001, 2005) or adaptive (Strange, Hejazi, Tang, 2006) mechanisms. These are now relatively well understood, though we still have little solid evidence as to their relative contribution and as to how they drive the geographic concentration of industries. But are these mechanisms *causes* of clustering? As I have shown before, it is tempting to believe so. First, there is a positive association between proxies for these mechanisms and the geographic proximity of industries. Second, simple simulation models of coagglomeration tells us 'yes, on average' (O'Sullivan and Strange, 2016). However, Section 2.1 also shows that coagglomeration can occur for reasons that do not reflect coagglomeration economies. Hence, we have to carefully address those various identification issues in any empirical application.

While coagglomeration patterns are prone to identification challenges, the problems are far less severe than in older approaches to the identification of agglomeration mechanisms. To see this, let me briefly go back to the first-generation models of identification. These models usually used a measure of geographic concentration of an industry — such as the Ellison-Glaeser index — and regressed it on what were considered reasonable proxies of the mechanisms (see, e.g., Rosenthal and Strange, 2001). There are three fundamental problems with that approach. First, working off a limited number of industries, we only have few observations. Second, the Ellison-Glaeser index is relatively time-persistent for the given spatial units, so panel techniques are hard to apply, which raises the issue of unobserved confounding factors. Third, and most importantly, there is a fundamental mismatch between the theoretical question and the data. Take, for example, the case of input-output links. Clearly, by definition input-output links operate *between two industries*. For some pairs ('Motor Vehicle Manufacturing' and 'Motor Vehicle Parts Manufacturing'), the links are strong, whereas for other pairs ('Motor Vehicle Manufacturing' and 'Textile Mills'), the links are weak. It is then hard to see the theoretical link between the geographic concentration of an industry and its 'overall reliance' on intermediate inputs as, e.g., measured by the ratio of intermediate input purchases to value added. Coagglomeration patterns, by allowing us to look more precisely at the channel predicted by theory, help us to exploit identifying variation across industry pairs and thus provide better identification of the theoretical mechanism. The idea underlying the second-generation models is hence to look at the link between economic proximity (mechanisms) and geographic proximity (coagglomeration). Since industries are always close to themselves (by definition) along (almost all) dimensions, there is not much useful information for identification there. Since industry pairs may be close in some dimensions, but distant in others, coagglomeration patterns provide more identifying variation, especially when the aim is to sort out the mechanisms.[32]

As I showed in Section 3.1, the data reveal that $c_{ij}^t = f\left(\mathbf{M}_{ij}^t\right)$. This suggests that we run regressions of the following form:

$$c_{ij}^t = \mathbf{M}_{ij}^t \alpha + \mathbf{X}_{ij}^t \beta + \xi_{(i,j)} + \delta_t + \epsilon_{ij}^t, \tag{2}$$

where c_{ij}^t is a measure of coagglomeration of industries i and j at time t; $\mathbf{M}_{ij}^t$ is a vector of 'mechanisms', and the α's are our main coefficients of interest; $\mathbf{X}_{ij}^t$ is a vector of (time-varying) industry-pair specific controls; and $\xi_{(i,j)}$ and δ_t are industry and year fixed effects.

Although second-generation models using coagglomeration patterns — such as (2) — have more identifying variation and flexibility to allow for the identification of the causal mechanisms of geographic concentration, they still have to deal with serious threats to identification. Industries can concentrate for many reasons that are unrelated to the mechanisms that we have in mind (omitted variables), and their concentration could cause the mechanisms to become stronger (reverse causality). There are also more subtle issues that I will discuss in what follows.

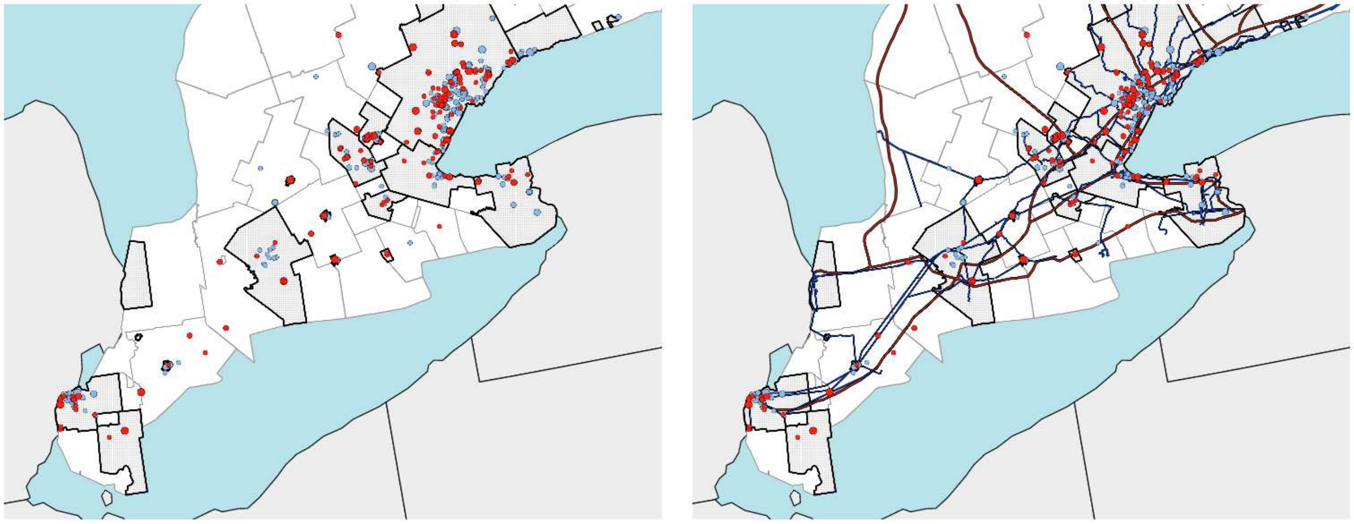
Threat 1. Spurious colocation. Natural advantage, unobserved local policies, and other factors that jointly target industries not related by coagglomeration economies may lead industries to cluster in the same place (see Section 2.1). As panel (a) of Figure 8 shows, although plants in 'Motor Vehicle Manufacturing' (NAICS 3361) and 'Motor Vehicle Parts Manufacturing' (NAICS 3363) seem to colocate systematically, they also all seem to colocate with major infrastructure, as panel (b) highlights. These potential confounding factors need to be controlled for in any serious analysis.

Building on work by Kim (1995) and Ellison and Glaeser (1997, 1999), the literature has developed methods to at least partly control for locational advantage. This is mainly done by constructing counterfactual agglomeration or coagglomeration patterns based on proxies for natural and locational advantages. Ellison and Glaeser (1999), for example, use 16 different proxies for production costs, labor force composition, and the presence of natural resources to estimate a model that interacts an industry's intensity of use in those proxies with a measure of local abundance of those proxies at the U.S. state level. This allows them to construct a

---

[32] Recent third-generation models allow to exploit even more identifying variation by looking at heterogeneous effects across industry pairs (Faggio et al., 2015) or functions within industry pairs (Behrens and Guillain, 2016). I will briefly return to these models that capture heterogeneous effects — and not average effects — in Section 4.

Figure 8: Locational advantage threatens the identification of agglomeration mechanisms.



(a) Colocation patterns.

(b) Major infrastructure.

baseline distribution of employment based on how sensitive industry location is with respect to those locational fundamentals. The counterfactual baseline distribution is then included as a control via $\mathbf{X}_{ij}^{t}$ in (2).

When microgeographic data are available, other techniques can be used. Klier and McMillen (2008) estimate, for the case of the U.S. automotive industry, a conditional logit model of the probability that an automotive plant chooses to locate in a specific county based on the county's characteristics. They then construct counterfactual distributions by randomly assigning plants to counties based on the logit probabilities. Behrens and Guillain (2016) uses a similar technique for the case of Canadian manufacturing. We estimate a probit model of the probability that a plant in some industry chooses to locate at a specific site, based on site characteristics such as distance to major infrastructure (highway, railway, sea port, airport), distance to U.S. industries and the border, local population, education of the local workforce, and a number of other site-specific variables constructed using GIS. Once industry-specific choice probabilities have been estimated for each site, plants in each industry are counterfactually allocated to their most likely sites, in decreasing order of probability with the oldest plants being allocated first. Using those counterfactual industry distributions, we then recompute coagglomeration measures between industries based on the counterfactual point patterns, and those measures are included as controls in the regression (2).

Finally, simple additional controls can be used to complement the analysis. The literature has, e.g., used an industry's shares of intermediates sourced from, and output sold to, primary industries. More recently, an industry's share of service intermediates has been considered. Note that the interactions with service industries have not been really analyzed until now, although some manufacturing activities rely significantly on external services (and cluster proponents rightly insist on the fact that we need to look jointly at manufacturing and services).

Services themselves are strongly agglomerated and some manufacturing industries may end up coagglomerated because they require proximity to the same services.

How important are locational fundamentals in practice? The bottom line is that locational advantage seems to explain 20%–30% of the variation in manufacturing (co)agglomeration patterns in Canada, and about as much in the agglomeration patterns of U.S. industries (Ellison and Glaeser, 1999). Although this means that locational advantage matters and needs to be controlled for, it seems to be only a small part of the overall (co)agglomeration story.

Threat 2. Omitted variables and reverse causality. There are many potentially unobserved factors that should be taken into account. Some industries are, for example, more urban, more capital intensive, and have larger plants than others. Industry fixed effects $\xi_{(i,j)}$ in (2) can deal with level differences in coagglomeration patterns of industries. Ideally, we would like to run real panel regressions, by including industry-pair fixed effects, but more aggregated measures of coagglomeration — like the Ellison-Glaeser index computed at a larger spatial scale — display little identifying variation across time. This makes panel regressions problematic. However, point-pattern based measures of coagglomeration at a small spatial scale remain more variable since there is a large amount of plant-level churning in the data. These measures thus allow potentially for the inclusion of industry-pair fixed effects, but more work is required here (see Behrens et al., 2015; Behrens and Brown, 2016).

Another threat to identification is the potential for reverse causality: some industries can colocate for reasons unrelated to interactions between them, and afterwards develop inter-industry links. Hence, input-output links, labor pooling, or knowledge sharing *may reflect coagglomeration patterns and not cause them*. For example, industries that end up being located close together by chance start buying from each other because of cost savings. Hence, the 'economic proximity' metrics may be endogenous in the sense that geographic proximity makes industries and plants more similar along various aspatial dimensions. This problem is obvious for the workforce metrics that are used in the literature: if plants end up in the same place, for whatever reason, they will hire from the same labor pool (since commuting patterns of workers are fairly localized).

The solutions adopted to date in the literature are classical instrumentation strategies. The most common approach is to rely on 'other country' instruments (see Ellison et al., 2010; Faggio et al., 2015; Behrens and Guillain, 2016), whereby the economic proximity of industry pairs in one country is instrumented by the economic proximity of that pair in another country. The other strategy used by Ellison et al. (2010) is that of spatial instruments, whereby the proximity metric is computed using data on plants in locations where no significant coagglomeration with the other industry occurs. While interesting, the latter approach is very data demanding since it requires, for example, detailed input-output measures at the plant level.

Threat 3. 'Third industry effects'. One problem that has not yet been seriously recognized in the literature is that *any geographic proximity relationship is transitive*: if plants in industries i and j are close (because they interact), and if plants in industries j and k are close (because they interact), then plants in i and k are close (but this does not mean that they necessarily interact). Spurious colocation may then take place independently of the underlying agglomeration mechanisms, triggered by the existence of industry j. This raises the problem of how to think about the coagglomeration of i and k. Is it spurious because of the existence of some 'third industry' j? To give a concrete example, let us look at Table 2. 'Motor Vehicle Manufacturing' is coagglomerated with 'Motor Vehicle Parts Manufacturing' for obvious reasons: these two industries are linked by one of the largest input-output coefficients in the Canadian economy. The latter industry includes subindustries like 'Motor Vehicle Seating and Interior Trim manufacturing', which may source inputs from 'Fabric Mills', 'Textile and Fabric Finishing and Fabric Coating', or 'Textile Furnishings Mills'. Although 'Motor Vehicle Manufacturing' does not directly source from the latter industries, they belong to the same production chain and are thus linked by a third industry. In the end, all may thus end up being close geographically.

The problem of third industry effects is a 'standard ij-problem', like in the gravity equation in migration or in international trade (Anderson and van Wincoop, 2003). The fundamental question is whether we can explain what is going on between the pair ij using solely variables pertaining to i and j? The answer clearly has to be 'no': the world is multilateral and not bilateral. As in a gravity equation, regressing a coagglomeration measure c_{ij}^t on variables that pertain only to industries i and j (and their interactions ij) misses the crucial general equilibrium structure of the economy. Industries are linked through complex networks. Note that this problem is totally *different from natural advantage* since it arises endogenously in the system through the joint location decisions of all industries.

To my knowledge, this problem has not been recognized, let alone solved in the literature. Ideally, we would like to have a structural model that tells us how to control for these effects. We do not have such a model to date. In Behrens and Guillain (2016) we try to address this issue by including in an ad hoc way the correlation coefficient of the coagglomeration measures of industries i and j with all other industries as a reduced-form control into the analysis. This control may be overly strong and pick up too much variation in the data (because a large set of 'urban' industries are strongly coagglomerated between them in the Canadian context at least). Alternatively, we can proceed by re-estimating the model by excluding industries that are generally too similar in terms of their overall location patterns with respect to third industries.[33] Preliminary findings suggest that controlling for third industry effects is important and influences the magnitude — but not the statistical significance — of the

---

[33] We could also consider controlling for an industry's position in the production chain, using e.g. an 'upstreamness measure' à la Antràs, Chor, Fally, and Hillberry (2012). However, that type of measure is probably relatively stable across time, so that the industry fixed effects should take care of it.

agglomeration mechanisms.

Threat 4. Within-plant agglomeration. The last threat I discuss is subtle and usually goes unnoticed due to data limitations. It is known that geographic concentration occurs among plants belonging to the same multiunit firms (see, e.g., Duranton and Overman, 2008), and that there are agglomeration economies *within firms* (see Alcácer and Delgado, 2016). What about complementarities that give rise to agglomerations *within plants*?

The question is one of the 'boundaries of the plant', and the fact that we usually do not observe all activities going on within plants. Most data sets report a plant's primary activity only. Yet, plants have often several activities. This may lead us to understate the extent of coagglomeration. Consider a plant in 'Motor Vehicle Manufacturing' (NAICS 3361) that also produces some of the intermediates it uses (i.e., it is active in 'Motor Vehicle Parts Manufacturing', NAICS 3363). Assume that I only see 3361 as the activity reported by the plant in my dataset. Then, I will underestimate the extent of coagglomeration c_{ij}^t of NAICS 3361 and 3363, because part of it is very close (at zero distance), but not reported in the data (since the unit of observation is the plant's primary activity). If this bias in c_{ij}^t is uncorrelated with the mechanisms $\mathbf{M}_{ij}^t$ in (2), we are fine. However, there is no reason for this to be so. The 'internalization decision' — which activities the plant performs jointly — is quite naturally correlated with those mechanisms.

Table 5: Within-plant agglomeration as threat to identification.

	within plant agglo.	input-output	knowledge flows	occ. similarity
input-output	**0.451***			
knowledge flows	**0.130***	0.098*		
occupational similarity	**0.390***	0.294*	0.104*	
labor movement	**0.491***	0.391*	0.123*	0.358*

Notes: * denotes significant at 1%. The table reports simple correlation coefficients between the different economic proximity measures. The results are based on 10,965 industry pairs (pooled for 2001, 2003, and 2005). The results are virtually identical for each individual cross-section, with all coefficient significant at 1%. Based on data from Behrens and Guillain (2016).

Table 5 shows that within-plant agglomeration — which corresponds to the most extreme form of spatial proximity — is strongly correlated with the mechanisms we have in mind. To assess this, I construct a measure of 'within-plant agglomeration' that reports the share of plants in industry i that also report secondary activities in industry j. As can be seen from the table, plants are more likely to operate in multiple industries that are: (i) strongly linked in input-output terms; (ii) draw on similar knowledge bases; and (iii) and occupy similar types of workers. These are all the agglomeration mechanisms we usually have in mind. Table 5 suggests that the coefficients we estimate for the mechanisms are likely to be downward biased. The extent of coagglomeration of industries with strong economic proximity is probably underestimated — since we focus on each plant's primary activity only — and the extent of underestimation is associated with the strength of the drivers. This is potentially a source of concern, but we have no clear idea how strong the bias is. And, in any case, this suggests that the coefficients we estimate for the determinants are probably conservative.

To summarize, while coagglomeration measures are very useful tools for identifying the causal mechanims driving the geographic clustering of industries — as compared to older and less direct approaches — they are not failsafe. Despite a number of more subtle unadressed identification issues, the literature using coagglomeration patterns has made substantial progress in the identification of mechanisms. Most agglomeration mechanisms survive the identification challenge and we have now causal evidence that input-output links and aspects linked to the similarity of workers and the mobility of workers across industries *cause industries to coagglomerate*. Evidence for knowledge spillovers — as measured by patent citations across industries — is much weaker. In Section 4, I return to this point, which is probably linked to both bad measurement (patent citations do not really capture knowledge spillovers) and the fact that knowledge flows are embodied in labor movements and input-output links. There, I will also briefly discuss the third-generation approaches that have started to further 'slice up' the identifying variation in the data to more cleanly identify the mechanisms and to account more fully for heterogeneity in those mechanisms across industries.

3.3 *Evaluating outcomes*

The most important policy question is: How does clustering, as perhaps induced by cluster policy, affect outcomes?[34] This is the area where we know the least, so some of the material I present here is brand new and necessarily remains to be developed more fully in the future. I focus on the links between clusters, that exist for whatever reasons, and outcomes. The literature has considered many outcomes of policy interest: (i) productivity and innovation (e.g., Duranton et al., 2010; Buzard et al., 2015); (ii) export participation and performance (Koenig, Mayneris, and Poncet, 2010; Ramos and Moral-Benito, 2013; Martin, Mayer, and Mayneris, 2013); and (iii) the resilience of clusters, or of firms in clusters, to adverse shocks (Martin et al., 2013; Delgado, Porter, and Stern, 2016b; and Behrens et al., 2016a,b).[35]

To understand the fundamental issues involved in assessing the *causal effect* of clusters on outcomes, consider the following simplified approach to the estimation of agglomeration economies, which I adapt from Combes and Gobillon (2015). Take outcome y for plant p, located in ℓ and at time t. To fix ideas, assume that y is productivity. Consider the

---

[34] I will not discuss here the scant evidence on the efficiency of cluster policy in actually attracting economic activity (see Duranton et al., 2010; Duranton, 2011; and Martin, Mayer, and Mayneris, 2011, for reviews and details). I take the existence of clusters as given.

[35] After 'clusters', 'resilience' has become the new buzzword. See Martin and Sunley (2015) for a critical discussion of the concept of resilience.

following two-step estimation procedure which is often used in the literature on agglomeration economies. In the first stage, we estimate location fixed effects $\widehat{\beta}_{\ell(p,t)}$ from the following linear regression:

$$y_{p,t} = u_p + \mathbf{X}_{p,t}\theta + \beta_{\ell(p,t)} + \epsilon_{p,t}, \tag{3}$$

where u_p is a plant fixed effect that captures time-invariant unobserved factors of the plant; $\mathbf{X}_{p,t}$ are time-varying plant characteristics; and $\beta_{\ell(p,t)}$ equals 1 if plant p is in location ℓ at time t, and 0 otherwise. The location fixed effects $\beta_{\ell(p,t)}$ capture unobserved location-specific factors that influence the productivity of plants in location ℓ — including all 'agglomeration economies' and locational fundamentals. As I show below, 'locations' could also be time varying. In the second stage, we regress these location fixed effects on a vector of proxies $\mathbf{Z}_\ell$ of mechanisms that we believe are relevant in generating agglomeration economies (e.g., the location's population size or density):

$$\widehat{\beta}_{\ell(p,t)} = \mathbf{Z}_\ell\gamma + \eta_\ell. \tag{4}$$

Note that we need to control for reverse causality in the second stage (4). Indeed, theory tells us that productive locations grow large since they attract firms and workers in spatial equilibrium (the 'endogenous quantity of labor' argument; Combes et al., 2008). The literature uses several instrumentation strategies for population size and density $\mathbf{Z}_\ell$. Commonly used instruments include: (i) historical instruments (lagged populations or density, first rail connection; Ciccone and Hall, 1996), which exploit persistence in historical urbanization patterns; (ii) geological instruments (soil type, slope, seismic hazard; Rosenthal and Strange, 2008; Combes, Duranton, Gobillon, and Roux, 2010), which influence the spatial equilibrium via shifts of the cost-of-living function $C(L_c)$; and (ii) geographical instruments (distance from eastern seaboard in the U.S. case; Ciccone and Hall, 1996), which exploit persistence in historical settlement patterns. The foregoing instrumentation strategies work fairly well (especially deeply lagged populations, the geological instruments being weaker). Moreover, the literature finds that reverse causality only seems to be a minor issue. In other words, OLS estimates of (4) provide a good approximation of the underlying causal relationships (see, e.g., Combes et al., 2011, for an extensive discussion).

It is tempting to adapt the foregoing analysis to the case of clusters. After all, I have been vague as to what constitutes a 'location', so a location could just be some appropriately identified geographic cluster. Let me thus now consider clusters, and try to apply the same approach. I simply take outcome y for plant p located in *cluster* c at time t, once clusters have been appropriately delineated (see Section 3.1). The first stage can be implemented as before:

$$y_{p,t} = u_p + \mathbf{X}_{p,t}\theta + \beta_{c(p,t),t} + \epsilon_{p,t}, \tag{5}$$

where $\beta_{c(p,t),t}$ now indicates whether plant p belongs to cluster c at time t.[36] In other words, we now estimate location fixed effects at the cluster level.[37] Note that not all plants need to belong to clusters, i.e., there can be an outside category for plants that are not in clusters, so that effects are measured relative to non-clustered plants.

Once the cluster fixed effects are estimated, the second stage can be thought of as before:

$$\widehat{\beta}_{c(p,t),t} = \mathbf{Z}_{c,t}\gamma + \eta_{c,t}, \tag{6}$$

where $\mathbf{Z}_{c,t}$ is a vector of proxies for the cluster's size. However, that second stage is now much more problematic for several reasons. In addition to the same problems discussed above, there are at least three additional issues to be addressed.

First, as explained before, clusters rarely conform to standard geographic units. Hence, we need to deal with non-standard geographies, which magnifies problems of omitted variables. For example, common cluster wage shocks cannot be adequately captured if wages are reported for spatial units that have no reasonable connection with the geographical or industrial extent of the clusters under consideration.

Second, reverse causality looms larger because of unobserved common shocks. As explained by Duranton et al. (2010, p.97), "[...] the increase (or decrease) of local employment may be, at least partly, due to conjunctural effects which also impact firms' performance. If there is a positive demand shock, the total factor productivity of firms (or workers' productivity) and the number of producing firms (or of workers employed by those firms) will both increase. There would be apparently a cluster effect whereas it is in fact only due to a conjunctural effect: that is the simultaneity bias. Researchers must account for those pitfalls carefully in order to highlight a real causal relationship between clusters and productivity." As an example, let me return to the automotive industry. During the financial crisis, 'Motor Vehicle Manufacturing' was hit by an important demand shock. As automobile demand contracts, so does demand for intermediates. Hence, there is a positive co-movement between the activity in 'Motor Vehicle Manufacturing' (NAICS 3361) and 'Motor Vehicle Parts Manufacturing' (NAICS 3363). This may be interpreted as evidence for the importance of cluster effects, even if it has no such interpretation at all in my example. The problem that y may drive $\mathbf{Z}$ via β seems especially severe for clusters, because clusters occur at relatively fine geographic and industrial scales. Also, many clusters are dominated by 'anchor firms' (Wolfe and Gertler, 2004) or large 'leaders'

---

[36] Note one key difference between (3) and (5): in the latter, the location fixed effect is indexed by t. Indeed, clusters change more than 'locations' (e.g., cities) through time, both in terms of industrial composition and in terms of geographic location. See Figure 7, which shows how substantially the geography of clusters can change over a short horizon. One 'solution' to this problem is to delimit clusters in a base year and to hold them constant afterwards. This allows one to keep the unit of analysis comparable across time but raises the issue of what we are really looking at after a couple of years if major changes take place.

[37] Industry is a criterion included in assigning a plant to a cluster, but we do not index this explicitly to save notation. Two plants p_1 and p_2 in the same place ℓ can (and generally will) be assigned to different clusters.

(Martin et al., 2013), both of which can — by their sheer size — drive cluster dynamics in the presence of unobserved idiosyncratic shocks.

Last, self-selection and sorting may be an issue. Better workers and more productive firms that grow faster might self-select into clusters. This should be controlled for when thinking about the productivity and wage effects of clustering. It needs to be addressed in any regression where the 'cluster' is on the right-hand side, and where the explained outcome is either productivity or wage.

What possibile solutions do we have? First, observe that standard instrumentation strategies will not work. The reason is that they are applicable to overall economic activity in location ℓ, but not specifically to clusters, which usually are a (small) subset of local economic activity. For example, 'only' 16% of Manhattan's employment was in the 'Finance and Insurance' industry in 2008, though this is clearly one of the big clusters that comes to mind.[38] How can we instrument 'Silicon Valley', or 'Wall Street', or the 'automotive cluster' in the Detroit-Windsor corridor? Silicon Valley 150 years ago does not seem to be very helpful. Historical instruments cannot work, since those that are valid are unlikely to be relevant and those that are relevant are unlikely to be valid. Geological instruments seem of equally limited use. We could use soil or slope characteristics of Manhattan, but why should these affect only affect the financial industry directly via supply of developable land or construction costs? They will affect overall economic activity, which affects the financial industry via a large variety of channels, rendering the instrument invalid.

This leaves us with 'solutions' that are relatively unsatisfying: dynamic panel techniques or GMM methods that rely on differenced and lagged variables as *internal instruments* (see Duranton et al., 2010, pp.110-111). These techniques are known to be problematic in a spatial context — where time variation of many variables is limited and where shocks are serially correlated — so we would like to at least check the robustness of the results against those obtained using *external instruments*. Here is where things get tricky, because as explained above we do not have good instruments. This is where coagglomeration patterns can be helpful.

Constructing coagglomeration instruments. Finding appropriate instruments for overall spatial activity is already hard. Finding industry-specific time-varying instruments that allow us to approximate clusters is even harder. Strange (2009, p. 12) observes that "[t]he processes in question operate at a fairly local level [...] so there are significant challenges to finding plausible instruments and disaggregated data [...] This means that agglomeration research continues to involve attempts to find better instruments or better disaggregated data." It is here that coagglomeration patterns may again prove helpful.

---

[38] See Edward L. Glaeser "Wall Street isn't enough", available online at `http://www.city-journal.org/html/wall-street-isn't-enough-13461.html`.

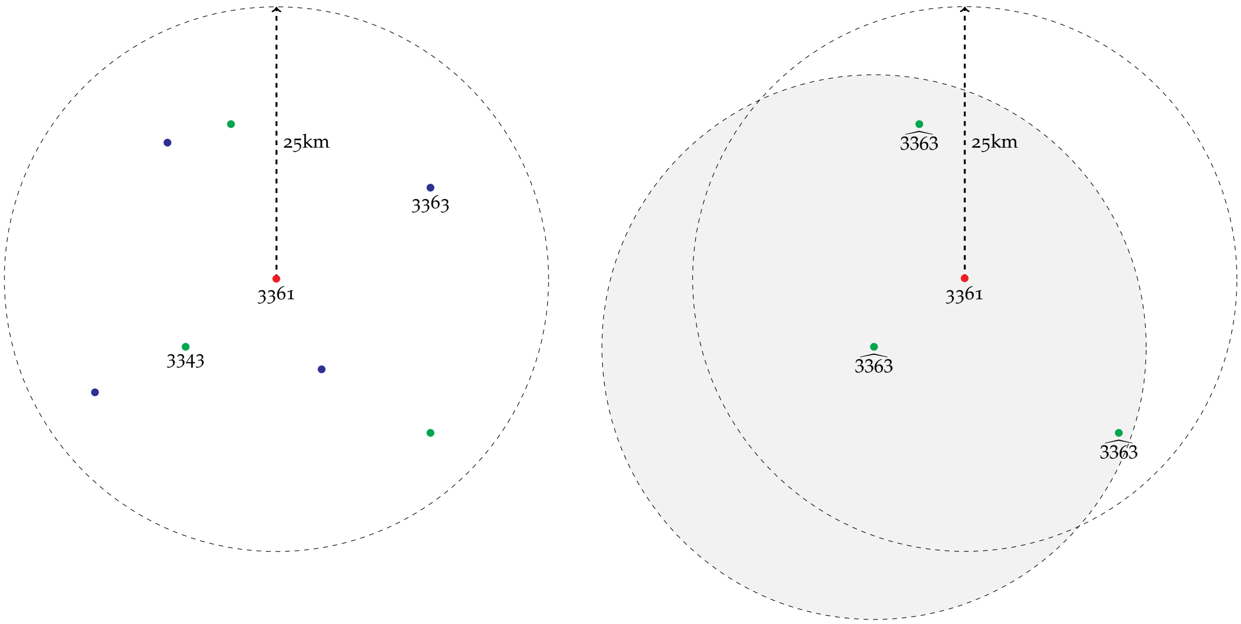
Assume that we want to instrument cluster employment $\mathbf{Z}_{c,t}$, which is the measure of the cluster's size that we think is a relevant proxy for the cluster's performance. Think about my automotive example and consider that $\mathbf{Z}_{c,t}$ is the employment in 'Motor Vehicle Parts Manufacturing' (NAICS 3363), which we want to use to evaluate productivity in 'Motor Vehicle Manufacturing' (NAICS 3361). We need to find industries that: (i) predict well this cluster employment $\mathbf{Z}_{c,t}$; and (ii) do not directly influence the outcome $y_{p,t}$. Those would be valid instruments to evaluate the effect of clusters on productivity.

We can construct such instruments from coagglomeration patterns. The basic idea is like in social networks identifying peer effects: we use both transitivity and intransitivity — my neighbors' neighbors or my friends' friends, provided they do not influence me directly (see Bramoullé et al., 2009, who refer to this as 'intransitive triads'). For the sake of concreteness, go back to my example. I want to assess the causal impact of more automobile parts suppliers on automobile manufacturers, i.e., look at the impact of more 'Motor Vehicle Parts Manufacturing' (NAICS 3363; $\mathbf{Z}_{c,t}$) on the size/productivity/export performance of local 'Motor Vehicle Manufacturing'(NAICS 3361; $y_{p,t}$). As argued before, the obvious problem is that demand shocks to automobile producers change the size of the cluster, depending on the strength of links and the magnitude of the shock. One solution could be to find industries that predict cluster employment of automobile parts suppliers but are not strongly linked to that industry and, therefore are unlikely to display co-movements with the shocks. Recall from Section 2.1 that coagglomeration may reflect locational fundamentals or historical lock-in, and so may not reflect real links between industries. Also, since we have various metrics $\mathbf{M}_{ij}^{t}$ that proxy the interactions between industries, we can pick industries that are coagglomerated but are *not* strongly related via either input-output links — in this case the presumed channel of endogeneity — or occupational similarity, or common knowledge bases, or other common factors.

In my example, the problem is clear: NAICS 3361 and 3363 are strongly linked via buyer-supplier relationships, which makes identification of causal effects difficult. The input and output shares between 3361-3363 in 2007 were 62.4% and 80.6%, respectively. However, there are several other industries that are coagglomerated with 3363 but display fairly small input-output shares. For example, take 'Video and Audio Equipment Manufacturing' (NAICS 3343). The input and output shares between 3363-3343 are only 1.6% and 1.9%, respectively; and the are only 0.1% and 7.7% between 3361-3343.

In my simple example, the location of NAICS 3343 is a good predictor of the location of NAICS 3363. So I can fix some distance (e.g., 25 kilometers) and count employment in NAICS 3343 within that distance around each plant in NAICS 3361. The instrument $\widehat{\mathbf{Z}}_{c(p,t)}$ is then the predicted employment in NAICS 3363, based on its *national coagglomeration pattern* with NAICS 3343 and the spatial distribution of NAICS 3343 around the industry NAICS 3361 of interest. Figure 9 illustrates this approach.

Figure 9: Example of coagglomeration instruments.



In my example, the correlation between the instrument $\widehat{\mathbf{Z}}_{c(p,t)}$ and local employment in NAICS 3363 at less than 25 kilometers from plants in NAICS 3361 is 0.52 for all plants in Canada. This suggests that the instrument is highly relevant. Observe that I use coagglomeration patterns and input-output shares measured at the national level to predict local outcomes. Hence, conditional on controls, and given the weak input-output shares with between 3363-3343, it is unlikely that idiosyncratic demand shocks for y_p in NAICS 3361 directly influence $\widehat{\mathbf{Z}}_c(p,t)$, thus making the instrument valid.

There are obviously many ways to create these coagglomeration instruments. But the basic idea is always the same: some part of the local employment or plant composition is interacted with national coagglomeration patterns and, eventually, national employment changes (see Behrens, Boualam, Martin, Mayneris, and Mion, *in progress*). It is worth noting that this approach turns the logic used to identify the mechanisms underlying geographic concentration on its head. To understand mechanisms, I use coagglomeration patterns and look at industries that have *strong links*. While these strong links are the researcher's boon to identify causal mechanisms, they are a bane when it comes to assessing outcomes (because they transmit idiosyncratic shocks). Hence, to work on outcomes, I use coagglomeration patterns and look at industries that have *weak links*. Theory tells us that these industry pairs exist — by spurious coagglomeration and transitivity — and they are easy to identify in the data.

Where do we stand? While there is clearly much more work to be done here, I believe that coagglomeration patterns will be useful in allowing us to better tackle identification issues concerning the causal effects of clusters on various economic outcomes. Coagglomeration patterns may help us address identification issues that cannot be addressed with standard instruments used in the agglomeration literature. Of course, when available, the clever exploitation of quasi-natural experiments may provide an alternative route (e.g., Greenstone, Hornbeck, and Moretti, 2010).

4. Open questions and frontiers

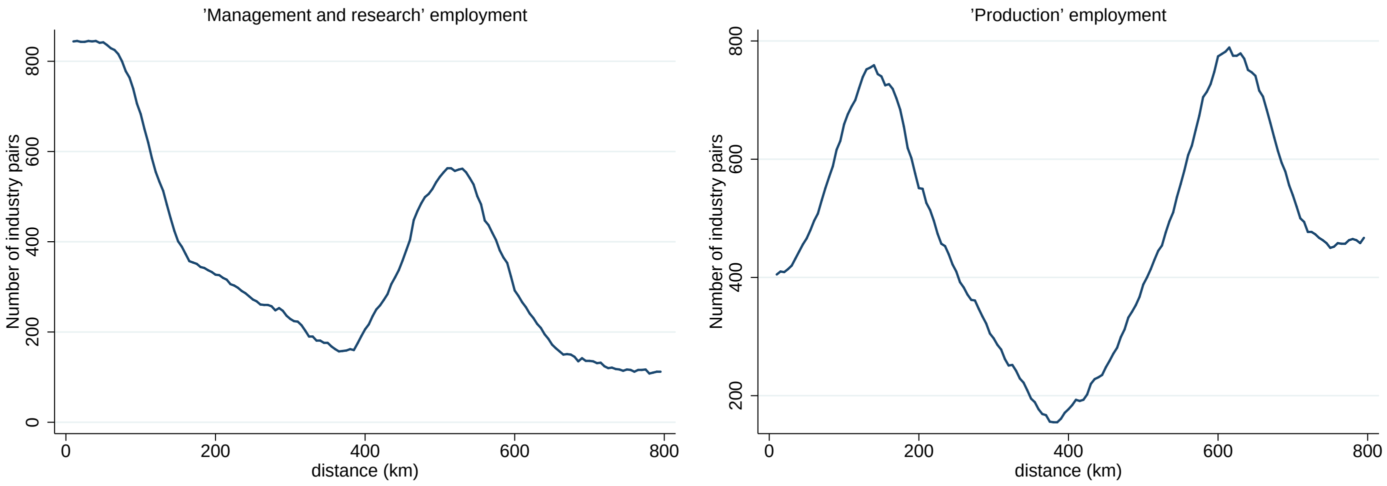
While much has been done these last few years, even more remains. I conclude this paper by discussing some of the open questions and by pointing out which frontiers we should aim to push back these next years.

Heterogeneous effects. Industries differ along many dimensions, and it is unlikely that they benefit in the same way from the different agglomeration channels. Even within industries, different plants serve different functions and are, therefore, again unlikely to depend in a unique way on the different agglomeration mechanisms. Acknowledging this heterogeneity is important in better identifying causal effects. The recent literature has moved beyond the second-generation models of coagglomeration to the estimation of heterogeneous effects using third-generation models. The idea is that different mechanisms affect industries differently, depending on industry characteristics (Faggio et al., 2015) or the functional distribution of employment (Duranton and Puga, 2001, 2005; Behrens and Guillain, 2016). Heterogeneity can occur along many informative dimensions: (i) industry and plant characteristics (Faggio, Silva, and Strange, 2015; Howard, Finn, and Tarp, 2015); (ii) occupational differences between workers (Gabe and Abel, 2016); and (iii) functional differences between plants (Behrens and Guillain, 2016). Another difference may lie in the types of clusters and regions (Martin, Mayer, and Mayneris, 2013; Delgado, Porter, and Stern, 2016b) that are part of the analysis.

As just stated, most recent third-generation models refine the analysis by exploiting additional identifying variation related to heterogeneity in coagglomeration patterns. Faggio et al. (2015), for example, estimate models by splitting industry pairs along various dimensions (industries with large vs small plants; young vs old plants; multiunit firms vs standalone plants). They document substantial heterogeneity between industries in the importance of the agglomeration mechanisms. Behrens and Guillain (2016) take a more 'functional approach', splitting overall employment — usually used to compute coagglomeration measures — along functional lines (e.g. Management and R&D employment, clerical employment, production employment). Figure 10 refines panel (b) of Figure 5 by splitting overall employment into separate 'Management and R&D' employment (panel (a) of Figure 10) and 'Production' employment (panel (b) of Figure 10).

As Figure 10 shows, 'Management and R&D' employment between industry pairs is more frequently and strongly coagglomerated at short geographic distances (with a much higher first peak), whereas 'Production' employment is more strongly colocated at intermediate distances

Figure 10: Heterogeneity in coagglomeration patterns by functions.


(a) Management and R&D. (b) Production.

(about 100 kilometers) and has a somewhat larger second peak. These patterns suggest that the colocation of management and R&D is more distance sensitive and tends to occur within the large metro areas, whereas production is more strongly coagglomerated at intermediate distances and between *different* large metro areas (thus the second peak). Note that these patterns are perfectly compatible with the 'functional specialization' à la Duranton and Puga (2001, 2005), where large cities host the (cross-industry) coagglomeration of higher-order functions, wheras lower-order functions are located more prominently in different more specialized cities. Exploiting these different patterns is likely to be useful to learn more about the different underlying determinants of geographic concentration, since we expect that functions that are intensive in some coagglomeration mechanism (e.g., input-output links for production, or knowledge exchange for management) will display geographic patterns compatible with those mechanisms.

Better proximity measures. As with all empirical work, the quality of the results on coagglomeration depends on the quality of the data used. While input-output links, workforce similarity of industries, and labor mobility across industries can be measured in a fairly satisfying way, the same does not hold true for knowledge spillovers. Yet, the latter have been among the most studied aspects in agglomeration empirics and have attracted the most attention from cluster proponents and policy makers. One of the general big challenges in the coming years will be to find better measures of pairwise industry interactions concerning 'knowledge flows'. Disentangling the parts of knowledge flows that are embodied in labor movements (e.g., Wolfe and Gertler, 2004) and the exchange of intermediate inputs (see Howard et al., 2015, who try to dissociate that aspect using firm-level survey data) is also an important and difficult task. Recent contributions have started to exploit more finely the geographic patterns of knowledge flows (Murata, Nakajima, Okamoto, and Tamura, 2014; Kerr and Kominers, 2015). Still, these analyses are based on patent-based measures of knowledge exchange. Going beyond these measures, which seem quite narrow in scope and of limited usefulness, seems important.[39]

Knowledge spillover proxies are not the only measures that require refinement. For example, even if input-output coefficients are of high quality, it is unclear what exactly they are capturing. As Combes and Gobillon (2015, p.336, my emphasis) note, those measures would need to be interacted with a measure of transport costs: "Overall, this strand of literature is an interesting effort to identify the mechanisms underlying agglomeration economies. Ultimately though, it is very difficult to give a clear interpretation of the results, and the conclusions are mostly descriptive. This is due to the weak links between estimated specifications and theoretical models. [. . .] one needs to assume that industry characteristics used as explanatory variables really capture the mechanisms they are meant to, and have additive linear effects, whereas this is not certain. For instance, according to theory, two industries sharing inputs have more incentive to colocate when trade costs for these inputs are large. In that perspective, *variables capturing input-output linkages should be caused to interact with a measure of trade costs, but this is not done in the literature*." More generally, the costs of trading goods have received fairly scant empirical attention, although they are a theoretical staple of most agglomeration models.

Service industries. The examples that I have considered all draw on coagglomeration patterns of manufacturing industries. Yet, those industries constitute now a rather small share of overall employment and economic activity. Hence, there is an urgent need to learn more about coagglomeration patterns of service industries, or of service and manufacturing industry pairs. Kolko (2010) is, to the best of my knowledge, one of the few contributions to look explicitly at the coagglomeration of service industries. I conjecture that the next few years will see more work on service industries as more microgeographic data become available and as improved computational procedures allow us to handle larger data sets (computing the Duranton-Overman measures on large data sets remains computationally a serious challenge). Note also that the coagglomeration of service industries — or the coagglomeration of service and manufacturing industries — will require us to think about a number of new problems, especially when it comes to causality. Some service industries follow population and will tend to appear coagglomerated with most other industries, thereby exacerbating the '3rd industry' problem. Selecting interesting and economically meaningful industries to look at coagglom-

---

[39] Knowledge transfer through patents may not be strongly affected by location, and with multi-plant companies it is often firm specific. It may be enough for firms to colocate their research groups. For example, in England, Coventry now only has the research groups of the automobile manufacturing firms — assembly has migrated elsewhere. I thank Dan Bernhardt for pointing this out. Behrens and Guillain (2016) use the 'patent intensity' of different functions to show that patents are more strongly linked to the coagglomeration of management and R&D than to the coagglomeration of production.

eration patterns — like finance, insurance, R&D, and other high-end producer services and manufacturing industries — will be a first challenge to confront.

Cluster dynamics. Dynamic aspects of agglomeration and coagglomeration patterns are mostly absent from the literature. How and why do these patterns change? How persistent are they? Location patterns might be inherited from the past but: (i) some of them still change rapidly, due to substantial plant-level churning (which makes microgeographic measures rather fast-changing); and (ii) this does not explain why agglomeration and coagglomeration measures correlate systematically with changes in the determinants of agglomeration, even over relatively short horizons (see, e.g., Behrens, Bougna, and Brown, 2016). Much more work is called for here. There are many interesting studies on the life cycles of clusters and their dynamics, yet little is still known. We do not know how clusters evolve, and we understand fairly badly why some can fail, even rapidly (see, e.g., Pouder and St. John, 1996, for theoretical elements of cluster evolution and failure). Yet, the quick demise of clusters, their resilience to shocks, and their capacity to adapt and to evolve are key issues for regional policy that we simply need to understand much better than we do now. Dumais, Ellison, and Glaeser (1997, 2002) are important exceptions that have looked at dynamics and changes in agglomeration patterns. Those contributions are almost twenty year old and still relevant, thus revealing the little progress we have made. Looking in more details at changes in coagglomeration patterns, or how entrants (see Rosenthal and Strange, 2003) or exiting firms are coagglomerated in expanding or shrinking industries, may prove very useful. It would allow to investigate location choices — and changes therein — based on current attributes rather than just the effects of history and sunkness, which contribute noise in many ways.

Extensions beyond plants. Most applications of geographic clusters were initially concerned with socio-demographic outcomes unrelated to plants (e.g., the localized prevalence of diseases in some population). We have made progress in developing measures and tools that allow us to look at the coagglomeration of firms, plants, and industries. In turn, those tools could be readily adapted to measure the coagglomeration of different characteristics unrelated to plants. For example, they could help us to better understand the colocation patterns of different socio-demographic or socio-economic characteristics. We could apply them to the issue of segregation *within race* by income, or to the coagglomeration of different worker qualifications with industries to learn more about the spatial mismatch of firms and workers. See Gabe and Abel (2016) for an application that goes beyond plants and employment and looks at the coagglomeration of occupations. Mele (2013) and Carrillo and Rothbaum (2016) provide recent microgeographic approaches that more systematically deal with local segregation patterns and their decomposition along various dimensions. There is a vast horizon of possibilities that just awaits to be explored.

Better 'basic' theory. In the end, what is missing is a better and more fleshed out theory of coagglomeration. This is important to better frame the research questions, and to tighten the (largely non-existent) link between our empirical measures and the underlying theory (see Kerr and Kominers, 2015; O'Sullivan and Strange, 2016). Little work has been done in that direction until now, but it is sorely needed. We should strive toward a more general theory of coagglomeration, where heterogeneous firms and workers in different industries sort across locations to form clusters based on the existence or absence of coagglomeration economies. While this is likely to be very hard, some interesting progress in this direction has been made recently in the literature (e.g., Davis and Dingel, 2015; Gaubert, 2016).

References

Albouy, David Y., Kristian Behrens, Frédéric Robert-Nicoud, and Nathan Seegert. 2016. "The optimal distribution of population across cities." *In progress*, Université du Québec à Montréal.

Alcácer, Juan, and Mercedes Delgado. 2016. "Spatial organization of firms and location choices through the value chain." *Management Science*, forthcoming. Available online at `http://dx.doi.org/10.1287/mnsc.2015.2308`.

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with gravitas: A solution to the border puzzle." *American Economic Review* 93**(1)**: 170–192.

Antras, Pol, Davin Chor, Thibault Fally, and Russell Hillberry. 2012. "Measuring the upstreamness of production and trade flows." *American Economic Review* 102**(3)**: 412–416.

Atalay, Enghin, Ali Hortaçsu, James Roberts, and Chad Syverson. 2011. "Network structure of production." *Proceedings of the National Academy of Sciences* 108**(13)**: 5199–5202.

Behrens, Kristian, Brahim Boualam, and Julien Martin. 2016a. "The resilience of the Canadian textile industries and clusters to shocks, 2001–2013." CIRANO Report #2016RP-05, available online at `http://www.cirano.qc.ca/files/publications/2016RP-05.pdf`.

Behrens, Kristian, Brahim Boualam, and Julien Martin. 2016b. "Are clusters resilient? Evidence from Canadian textile industries." *In progress,* Université du Québec à Montréal.

Behrens, Kristian, Brahim Boualam, Julien Martin, Florian Mayneris, and Giordano Mion. 2016. "Coagglomeration instruments." *In progress*, Université du Québec à Montréal.

Behrens, Kristian, and Théophile Bougna, 2015. "An anatomy of the geographical concentration of Canadian manufacturing industries." *Regional Science and Urban Economics* 51**(C)**: 47–69.

Behrens, Kristian, Théophile Bougna, and W. Mark Brown. 2015. "The world is not yet flat: Transport costs matter!" CEPR Discussion Paper #10356, Centre for Economic Policy Research, London, UK.

Behrens, Kristian, and W. Mark Brown. 2016. "Transport costs, plant location, and geographic concentration: Evidence from Canada." *In progress.*

Behrens, Kristian, and Rachel Guillain. 2016. "The determinants of coagglomeration: Evidence from functional employment patterns." *In progress.*

Behrens, Kristian, and Frédéric L. Robert-Nicoud. 2015. "Agglomeration theory with heterogeneous agents." In: Duranton, Gilles, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics, vol.5*. North-Holland: Elsevier B.V., pp. 171–246.

Bernard, Andrew B., Andreas Moxnes, and Yukiko Saito. 2015. "Production networks, geography, and firm performance." CEPR Discussion Paper #10551, Centre for Economic Policy Research, London, UK.

Bleakly, Hoyt, and Jeffrey Lin. 2012. "Portage and path dependence." *Quarterly Journal of Economics* 127**(2)**: 587–644.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of Econometrics* 150**(1)**: 41–55.

Buzard, Kristy, Gerald A. Carlino, Hunt, R. M., Carr, J., and Tony E. Smith. 2015. "Localized knowledge spillovers: Evidence from the agglomeration of American R&D labs and patent data." Working Papers #15-3, Federal Reserve Bank of Philadelphia.

Carrillo, Paul E., and Jonathan L. Rothbaum. 2016. "Counterfactual spatial distributions." *Journal of Regional Science*, forthcoming (DOI: `10.1111/jors.12277`).

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

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2011. "The identification of agglomeration economies." *Journal of Economic Geography* 11**(2)**: 253–266.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial wage disparities: Sorting matters!" *Journal of Urban Economics* 63**(2)**:723–742.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2011. "Estimating agglomeration economies with history, geology, and worker effects." In: Glaeser, Edward L. (ed.) *Agglomeration Economics*. Cambridge, MA: National Bureau of Economic Research, Inc., pp. 15–66.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The empirics of agglomeration economies." In: Duranton, Gilles, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics, vol. 5*. North-Holland: Elsevier B.V., pp. 247–348.

Cressie, Noel A. C. 1993. *Statistics for Spatial Data*. John Wiley, New York, NY.

Davis, Donald R., and Jonathan I. Dingel. 2015. "The comparative advantage of cities." *Mimeo*, Columbia University.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. 2016a. "Defining clusters of related industries." *Journal of Economic Geography* 16**(1)**: 1–38.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. 2016b. "Clusters and the great recession." *Mimeo*, Temple University and Harvard Business School.

Diggle, Peter. J. 2003. *Statistical Analysis of Spatial Point Patterns.* Oxford University Press, New York, NY.

Dumais, Guy, Glenn D. Ellison, and Edward L. Glaeser. 1997. "Geographic concentration as a dynamic process." NBER Working Papers #6270, National Bureau of Economic Research, Inc.

Dumais, Guy, Glenn D. Ellison, and Edward L. Glaeser. 2002. "Geographic concentration as a dynamic process." *Review of Economics and Statistics* 84**(2)**: 193–204.

Duranton, Gilles. 2011. "California Dreamin': The Feeble Case for Cluster Policies." *Review of Economic Analysis* 3**(1)**: 3–45.

Duranton, Gilles, and Henry G. Overman. 2008. "Exploring the detailed location patterns of U.K. manufacturing industries using microgeographic data." *Journal of Regional Science* 48**(1)**: 213–243.

Duranton, Gilles, and Henry G. Overman. 2005. "Testing for localization using micro-geographic data." *Review of Economic Studies* 72**(4)**: 1077–1106.

Duranton and Puga, 2015. "Urban land use." In: Duranton, Gilles, J. Vernon Henderson, and William C. Strange (eds.) *Handbook of Regional and Urban Economics, vol.5*. North-Holland: Elsevier B.V., pp. 467–560.

Duranton, Gilles, and Diego Puga. 2005. "From sectoral to functional urban specialisation." *Journal of Urban Economics* 57**(2)**: 343–370.

Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of urban agglomeration economies." In: Henderson, J. Vernon, and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics, vol. 4.* North-Holland: Elsevier B.V., pp. 2063–2117.

Duranton, Gilles, and Diego Puga. 2001. "Nursery cities: Urban diversity, process innovation, and the life cycle of products." *American Economic Review* 91**(5)**: 1454–1477.

Duranton, Gilles, Philippe Martin, Thierry Mayer, and Florian Mayneris. 2010. *The Economics of Clusters.* CEPREMAP and Oxford University Press Inc., New York

Ellison, Glenn D., and Edward L. Glaeser. 1999. "The geographic concentration of industry: Does natural advantage explain agglomeration?" *American Economic Review* 89**(2)**: 311–316.

Ellison, Glenn D., and Edward L. Glaeser. 1997. "Geographic concentration in U.S. manufacturing industries: A dartboard approach." *Journal of Political Economy* 105**(5)**: 889–927.

Ellison, Glenn D., Edward L. Glaeser, and William R. Kerr. 2010. "What causes industry agglomeration? Evidence from coagglomeration patterns." *American Economic Review* 100**(3)**: 1195–1213.

Faggio, Giulia, Olmo Silva, and William C. Strange. 2014. “Heterogeneous agglomeration.” SERC Discussion Paper #152, Spatial Economic Research Center, *London School of Economics*, UK.

Fujita, Masahisa. 1988. *Urban Economic Theory.* Cambridge University Press. Cambridge, UK.

Fujita, Masahisa Paul R. Krugman, and Anthony J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade.* MIT Press. Cambridge, MA.

Fujita, Masahisa, and Jacques-François Thisse. 2002. *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge University Press, Cambridge, UK.

Gabe, Todd M., and Jaison R. Abel. 2016. “Shared knowledge and the coagglomeration of occupations.” *Regional Studies* 50**(8)**: 1360–1373.

Gaubert, Cécile. 2016. “Firm sorting and agglomeration.” *Mimeo*, University of California, Berkeley.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. “Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings.” *Journal of Political Economy* 118**(3)**: 536–598.

Helsley, Robert W., and William C. Strange. 2014. “Coagglomeration, clusters, and the scale and composition of cities.” *Journal of Political Economy* 122**(5)**: 1064–1093.

Henderson, J. Vernon. 2003. “Marshall’s scale economies.” *Journal of Urban Economics* 53**(1)**: 1–28.

Henderson, J. Vernon. 1974. “The sizes and types of cities.” *American Economic Review* 64**(4)**: 640–656.

Henderson, J. Vernon, Ari Kuncoro, and Matthew E. Turner. 1995. “Industrial development in cities.” *Journal of Political Economy* 103**(5)**: 1067–1090.

Howard, Emma, Carol Newman, and Finn Tarp. 2015. “Measuring industry coagglomeration and identifying the driving forces.” *Journal of Economic Geography*, forthcoming (`doi:10.1093/jeg/lbv037`).

Kerr, William R., and Scott D. Kominers. 2015. “Agglomerative forces and cluster shapes.” *Review of Economics and Statistics* 97**(4)**: 877–899.

Kim, Sukkoo. 1995. “Expansion of markets and the geographic distribution of economic activities: The trends in U. S. regional manufacturing structure, 1860–1987.” *Quarterly Journal of Economics* 110**(4)**: 881–908.

Klier, Thomas, and Daniel P. McMillen. 2008. “Evolving agglomeration in the U.S. auto supplier industry.” *Journal of Regional Science* 48**(1)**: 245–267.

Koenig, Pamina, Florian Mayneris, and Sandra Poncet. 2010. “Local export spillovers in France.” *European Economic Review* 54**(4)**: 622–641.

Kolko, Jed. 2010. "Urbanization, agglomeration, and the coagglomeration of service industries." In: Glaeser, Edward L. (ed.), *Agglomeration Economics*. NBER Books, University of Chicago Press, pp. 151–180.

Krugman, Paul R. 1991. *Geography and Trade.* MIT Press, Cambridge, MA.

Martin, Philippe, Thierry Mayer, and Florian Mayneris. 2013. "Are clusters more resilient in crises? Evidence from French exporters in 2008-2009." CEPR Discussion Papers #9667, Centre for Economic Policy Research, London, UK.

Martin, Philippe, Thierry Mayer, and Florian Mayneris. 2011. "Public support to clusters: A firm level study of French 'Local Productive Systems'." *Regional Science and Urban Economics* 41**(2)**: 108–123.

Marcon, Eric, and Florence Puech. 2010. "Measures of the geographic concentration of industries: Improving distance-based methods." *Journal of Economic Geography* 10**(5)**: 745–762.

Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.

Martin, Ronald L., and Peter Sunley. 2015. "On the notion of regional economic resilience: Conceptualization and explanation." *Journal of Economic Geography* 15**(1)**: 1–42.

Martin, Ronald L., and Peter Sunley. 2003. "Deconstructing clusters: Chaotic concept or policy panacea?" *Journal of Economic Geography* 3**(1)**: 5–35.

Mele, Angelo. 2013. "Poisson indices of segregation." *Regional Science and Urban Economics* 43**(1)**: 65–85.

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

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura. 2014. "Localized knowledge spillovers and patent citations: A distance-based approach." *Review of Economics and Statistics* 96**(5)**: 967–985.

O'Sullivan and Strange. 2016. "Coagglomeration." *In progress*, University of Toronto.

Otazawa, Toshimori, and Jos van Ommeren. 2015. "Inter-firm transaction networks and location in a city." Mimeo, *Kobe University* and *Vrije Universitaet Amsterdam*.

Page, Scott E., and Troy Tassier. 2007. "Why chains beget chains: An ecological model of firm entry and exit and the evolution of market similarity." *Journal of Economic Dynamics and Control* 31**(10)**: 3427–3458.

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

Porter, Michael E. 1990. *The Competitive Advantage of Nations.* New York, NY: Free Press.

Pouder, Richard, and Caron H. St. John. 1996. "Hot spots and blind spots: Geographical clusters of firms and innovation." *Academy of Management Review* 21**(4)**: 1192–1225.

Ramos, Roberto, and Enrique Moral-Benito. 2013. "Agglomeration matters for trade." Working Papers #1316, Banco de España.

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

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the nature and sources of agglomeration economies." In: J. Vernon Henderson, and Jacques-François Thisse (eds.), *Handbook of Regional and Urban Economics, vol.4*. North-Holland: Elsevier B.V., pp. 2119–2172.

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

Rosenthal, Stuart S., and William C. Strange. 2001. "The determinants of agglomeration." *Journal of Urban Economics* 50**(2)**: 191–229.

Strange, William C. 2009. "Agglomeration research in the age of disaggregation." *Canadian Journal of Economics* 42**(1)**: 1–27.

Strange, William C., Walid Hejazi, and Jianmin Tang. 2006. "The uncertain city: Competitive instability, skills, innovation and the strategy of agglomeration." *Journal of Urban Economics* 59**(3)**: 331–351.

Trefler, Daniel. 2008. "Canadian policies for broad-based prosperity." *Canadian Journal of Economics* 41**(4)**: 1156–1184.

Wolfe, David A., and Meric S. Gertler. 2004. "Clusters from the inside and out: Local dynamics and global linkages." *Urban Studies* 41**(5/6)**: 1071–1093.