Agglomeration Research in the Age of Disaggregation

William C. Strange
Rotman School of Management
105 St. George St.
University of Toronto
Toronto, ON M5S 3E6
Canada
wstrange@rotman.utoronto.ca
(416)978-1949

September 25, 2008

Abstract

This paper selectively reviews recent research on the economics of agglomeration. Theory offers a long list of forces that might explain the spatial concentration of economic activity. The increased availability of disaggregated data -- by firm, by worker, and by geographical unit -- presents opportunities to substantially advance the understanding of urban growth and industrial clustering. At present, however, there remains great uncertainty about the forces that produce agglomeration. This suggests that public policy towards cities ought to be characterized by caution.

JEL Code: R0

1. Introduction

Urban economics begins as a field with the observation that economic activities are unevenly distributed across places. This unevenness is obvious to anyone who has ever traveled by airplane. Cities give way to suburbs and to farming areas and then to places that are almost uninhabited. Overall economic activity is thus visibly uneven across space. So too are individual activities. Some cities are notable for tall buildings, the physical manifestations of business service activities. Other cities are notable for factories – metaphorical and literal smokestacks – the physical manifestations of manufacturing activities.

This unevenness is a kind of aggregation across space, and the perspective with which I have described it is aggregate as well. The eyes of fliers see large patterns, but distance hides the finer details of spatial unevenness. The perceived aggregation is not, of course, simply some misperception borne of travel fatigue. They appear plainly in the data. The populations of Toronto, Montreal, Vancouver, Ottawa, Calgary, and Edmonton (Canada's Census Metropolitan Areas -- CMAs -- with populations more than 1 million) are 45% of national population. These cities occupy 0.37% of Canada's land. And there is intra-city agglomeration too. This can be seen clearly in differences in building heights within cities. It can be seen only slightly more abstractly in differences in densities within metropolitan areas. For instance, the population density in the City of Toronto is 3939 people per square kilometer, while in the entire CMA it is 630. Very similar pictures emerge looking at data from North America (Holmes-Stevens (2004)), Europe (Combes and Overman (2004)), and Asia (Fujita et al (2004)).

The degree to which industries cluster -- another sort of agglomeration -- is also striking. This agglomeration is illustrated in Figure 1. The map describe the allocation of employment in furniture manufacturing at the county level in the Eastern U.S. One reason that this industry might exhibit spatial concentration is that its material inputs also exhibit spatial concentration, so there might be a tendency towards spatial concentration near input sources. Figure 1 is not consistent with an especially strong role for this sort of natural advantage. North Carolina has roughly the same amount of establishments in forestry as Maine or as Alabama (County Business Patterns data from 2005), and yet these states do not have anywhere near the employment in furniture. Some other agglomeration forces must be present. Furthermore, these forces must operate not just at the national level, but also at the regional and local levels. It is also clear from Figure 1 that within the dominant region for furniture manufacture there is also unevenness, with more concentration in the vicinity of High Point, North Carolina. So the agglomeration forces are localized geographically.

Although it is unsurprising, I need to observe that this sort of industrial clustering is not unique to the furniture industry. In fact, I picked this industry as an example because it is somewhat less well-known than other exemplars of industrial clustering: the Silicon Valley, the North Carolina Research Triangle, Motor City. It is worth pointing out that many of the popular labels for industrial clustering are streets: Bay Street, Wall Street, Rodeo Drive, Sand Hill Road. This again speaks to the localized nature of agglomeration.

This agglomeration of overall activity and of individual activities is, of course, the outcome of disaggregate individual decisions. This is reflected in two lines of economic research. The first deals with the microfoundations of agglomeration economies, while the second concerns the estimation of agglomeration effects using micro-data. I use the latter term broadly, to include data that is firm-specific or individual-specific or place-specific at a level of geography below the city level. These two lines of research are obviously closely related to each other, since theories of agglomeration effects generate predictions that can be taken to everimproving data. In discussing the two lines of research, I will be highly selective. More complete surveys of these lines of research can be found in Fujita and Thisse (2002), Duranton and Puga (2004), and Rosenthal and Strange (2004).

The paper has three parts. Section 2 considers microfoundations. It does so using a model sufficiently abstract that it can capture many of the ways that disaggregate individual decisions produce aggregate agglomeration. Section 3 considers representative empirical work making use of data that are disaggregated by firm, worker, or geographical unit. Section 4 concludes by briefly considering what economic research on agglomeration has to offer to policy debates on urban policy.

2. Microfoundations of agglomeration economies

2.1. Preliminaries

The question of why agglomeration occurs is an old one. Most literature reviews begin with Marshall's (1890) trinity: labor market pooling, input sharing, and knowledge spillovers. With regard to labor market pooling, Marshall discusses the ability of a large market to offer "a constant market for skill," where both employers and workers benefit from the imperfect correlations between market variables. In modern terms, a programmer at Apple might move to another Silicon Valley employer if Apple's fortunes were to decline. If Apple were to prosper, however, there would be nearby workers with appropriate skills who might be attracted. With regard to input sharing, the proximity of firms in related industries can support a mutually helpful market of upstream input suppliers. Los Angeles, for example, supports a range of activities that are inputs in film and television production. Similarly, the Silicon Valley's computer manufacturers, including Apple, clearly benefitted from close proximity to microchip producers. With regard to knowledge spillovers, Marshall argues that spatial concentration results in a situation where the "secrets of the trade, are, as it were, in the air." This applies the Steve Jobs visit to Xerox PARC that led to Apple's bringing to market such innovations as the mouse and its graphical user interface. It applies equally well to the geographic spread of financial innovation in cities such as Chicago, New York, and London.

As important as Marshall's work in this area has been, it is worth noting that some similar ideas are found in earlier work by Smith (1776; 2008) and von Thunen (1863; 1966), the latter a point made by Fujita (2000). The modern study of agglomeration begins with the work of Vernon (1960) and others as part of the New York Metropolitan Region Project. See Quigley (1998). To this, I would be remiss if I did not add Jacobs' books (1961, 1969): *The Death and Life of Great American Cities, The Economy of Cities*. The latter is more relevant here, although it clearly continues a line of work begun with the first book. Jacobs (1969) stresses diversity as a source of a city's strength.

2.2. Formal analysis of microfoundations

The work described above proceeded for the most part without mathematically developed economic theory. The theoretical economics literature on the microfoundations of agglomeration economies began to take shape in the 1980s, although there is certainly important prior work, such as Beckman's (1976) classic. One direction that the literature took in the 1980s (see Ogawa-Fujita (1980), Fujita-Ogawa (1982), and Helsley (1990)) was to consider intra-city agglomeration processes by relaxing the classic assumption of a pre-determined city center to which businesses and commuters are oriented by supposing a spatially decaying agglomeration benefit. These papers were concerned with the implications of agglomeration forces on urban spatial structure, but they were not primarily concerned with the forces responsible for the attraction of agglomeration. The earlier literature also devoted some attention to inter-city general equilibrium (Henderson (1974)).

The best-known line within the theoretical microfoundations literature is New Economic Geography (NEG). This approach generates agglomeration in monopolistic competition models of input-output linkages. Fujita (1988), Abdel-Rahman and Fujita (1990), Krugman (1991), and Ottaviano-Tabuchi-Thisse (2002) are some of the key contributions. The NEG literature remains very active. Bosker et al (2007), Cordoba (2008), Sato (2007), Cavailhès et al (2007), Murata (2007), Pflüger -Südekum (2008), Robert-Nicoud (2008) and Zeng (2008) are recent contributions. The key element of the NEG analysis is that the backward and forward linkages between producers and markets produce a natural tendency to agglomerate. This leads to the possibility of a core-periphery allocation, exactly the sort of uneven spatial allocation discussed in the Introduction. NEG models exhibit path dependent dynamics, with history mattering. Furthermore, NEG models also admit the possibility of catastrophic change. Specifically, small changes in transportation cost can produce discontinuous changes in the nature of the equilibrium economic geography.

There are many alternatives to NEG. As per Marshall (1890), agglomeration can reduce risk. The relationship between risk and agglomeration has been formalized by Krugman (1991) and by Duranton-Puga (2004). Risk also plays a role in recent papers by Combes and Duranton (2006), Strange et al (2006) and Almazan et al (2007). Also as per Marshall, agglomeration allows knowledge to spill over. Although there has been considerable empirical work on knowledge spillovers, there has been much less theoretical work. Glaeser (1999) presents a dynamic model of learning as an exogenous outcome of meetings between a city's agents.

Helsley and Strange (2004) present a model where the productivity of a given meeting is endogenous. Duranton and Puga (2001) present a model where cities are superior places for producing prototypes, given the Vernon-type instability of such products, but not necessarily for routine production runs. More generally, Arnott (1979) shows how agglomeration economies can be conceived of as public goods, an approach employed in Helsley and Strange (1994, 1997) among others.

The microfoundation that I can write about with greatest familiarity is matching (Helsley-Strange (1990, 1991). The key to this sort of model is that agents are horizontally differentiated and must find suitable match partners. The match may be conceived of as being between workers, between workers and firms, between upstream and downstream firms, or between ideas and actualizations. It thus captures aspects of Marshall's trinity, as well as the stability-uncertainty-diversity aspect of agglomeration emphasized separately by Vernon (1960) and Jacobs (1969). The next section will present a simple matching model more formally. There can be no question that the model fails to span the universe of possible microfoundations, but it does encompass a broad range of them. In order to maximize its coverage of different microfoundations, the model will be quite abstract in its formulation.

2.3. A simple matching model

In the model, there are two types of agents who match. On one side are workers. On the other are specialized input suppliers, knowledge workers, and employers requiring specialized skills. These match partners produce factors that are complementary to a worker's idiosyncratic attributes, provide specific complementary knowledge, or make highly specific uses of the worker's specific attributes. These agents will be thus referred to as "complements."

The novel feature of this model is that instead of supposing there to be two sides of the market that match (as with males and females), I suppose that the complements arise entrepreneurially from the population (as with hermaphrodites). For instance, entrepreneurs arise from the population of workers, and then proceed to hire from the same population.[1] Formally, I conceive of matching as a two-stage process. In the first stage, complements arise from the population of workers. This ensures that they earn a reservation level of utility (to be thought of as a kind of zero-profit condition). The marginal entrepreneur must, for example, achieve the same level of utility as an entrepreneur and as a worker. The level of utility will depend on the expected benefits from matching and the congestion of fixed factors associated with agglomeration. These are described in more detail below. In the second stage, matching takes place. There are many ways that this could be modeled. The objective here is to specify a process where a larger city has a thicker market and in some sense expands the possibilities for matching. I therefore suppose that each worker is matched with the complement nearest to it. More concretely, an entrepreneur hires the workers best suited to his or her activity. The value of this match to the participants and also the value of alternate matches will determine the equilibrium wage for workers and rewards for complements. I will not consider here the organization of workers and complements into firms. See Helsley-Strange (2007) and Egger and Egger (2007).

The payoffs of the workers and complements are determined as follows. All agents are assumed to be risk-neutral with addresses on the unit circle describing their specific characteristics. Let x be the address on the unit circle characterizing the complement's specialized input production, knowledge, or labor demand. Let y be the address, also on the unit circle, describing a worker matched with the complement. The distance between x and y, d = |x -

y|, then measures the quality of the match. If x = y, then there is a perfect match. If the match is not perfect, the complement incurs adjustment costs that increase with the distance in the characteristic space. Specifically, we suppose that the adjustment costs are linear, equal to a constant b times d. If the potential value of the match is denoted by a, then the gross value of the match is a - bd.

At the matching stage, there are m complements and n workers. I assume that neither workers nor complements knows either their own nor any other agent's location on the characteristic space. Some assumption like this is required for the matching to be nontrivial. In order to simplify the algebra of matching, I suppose that after becoming active, the m complement suppliers are spread evenly within a city, with addresses given by $x \in \{0, 1/m, 2/m,\ldots,(m-1)/m\}$ on the unit circle. Assuming that they are draws from some distribution adds to the complication of the model but not to its insightfulness. I suppose that the addresses of the workers are draws from the uniform distribution.

The wage paid to a worker is assumed to be the outcome of a bargain between the worker and the complement. This will depend on the degree to which different complements are well-suited to the worker's specific skills. Let $d_{(1)}$ denote the worker's distance from the nearest complement. Let $d_{(2)}$ denote the distance to the second-nearest. The worker can be valuably matched to either of these. It is natural to suppose that the result of the bargain between the worker and the complements is to split-the-difference between the value of the nearest and next nearest complements to the worker. This means that the wage paid by a worker by the nearest complement will be

$$w = a - b[d_{(2)} + d_{(1)}]/2. \tag{1}$$

The complement will receive the residual,

$$r = a - bd_{(1)} - w = b[d_{(2)} - d_{(1)}]/2. \qquad (2)$$

The assumption of even spacing of complements means that $d_{(2)} + d_{(1)} = 1/m$, which in turn implies that $d_{(2)} = 1/m - d_{(1)}$. Substitution into (1) and (2) gives

$$w = a - b/(2m), \qquad (3)$$

and

$$r = b[1/m - 2d_{(1)}]/2. \qquad (4)$$

In this setup, the wage is independent of a worker's exact location between two complements. A closer worker would be more productive (lower adjustment cost), but would have less bargaining power. These cancel each other out exactly in this situation. Since $\partial w/\partial m > 0$, wage increases in the number of complements. The residual that accrues to the complement is not independent of the worker's location. A closer worker is good for the complement both because more value is created and because more of it accrues to the complement in the bargain. Since $\partial r/\partial m < 0$, the residual decreases with the number of complements.

A complement will be matched with those workers closer to it than to any other complement. A particular complement's total payoff R will be the sum of the residual from these matches. This completes the characterization of the equilibrium matching in the second stage of the model.

In the first stage, complements choose whether or not to become active. There are two costs incurred when a complement becomes active, a resource cost F and the opportunity cost of the expected wage that the complement would have earned as a worker. The latter is given by (3). A complement's revenues are the sum of residuals from matches with workers that are closest to it. With m complements and n workers, the complement expects to have n/m matches. Given equal spacing of complements, $d_{(1)}$ will range from [0,1/(2m)]. The assumption of worker addresses arising from uniform draws on the unit circle implies that $d_{(1)}$ is also drawn from a uniform distribution, in this case on [0,1/(2m)]. This then implies that $E(d_{(1)}) = 1/(4m)$. Substituting this into (4) gives the expected residual accruing to a complement from a match,

$$E(r) = b/4m. \tag{5}$$

A complement's expected payoffs equal the expected number of matches multiplied by the expected value to the complement less the costs of becoming active. With n workers and m complements, the expected number of matches per complement is n/m. Given this, (3), and (5), the complement's expected total payoff, E(R), is given by

$$E(R) = bn/(4m^2) - F - (a-b/(2m)). \tag{7}$$

The equilibrium amount of complement activity will be a root of E(R) = 0. The relevant root is the positive one,

$$m(n) = (1/(F+a)) * [b/4 + [b^2/4 + bn(F+a)]^{1/2}]. \qquad (8)$$

The relationship of the number of complements is of primary interest here:

$$dm/dn = 1/2 * [b^2/4 + bn(F+a)]^{-1/2} > 0. \qquad (9)$$

An increase in population encourages complements to become active. It is this relationship between workers and complements that generates the model's agglomeration economy.

Of course, agglomeration is costly. Land is the most important of the many fixed factors that are congested as a city's population grows. Access to clean air and water are also congestible. In the short run, at least, so are roads and other public goods. I capture all of these with the increasing and convex cost function c(n), giving the cost incurred by an individual worker or complement at a population of workers equal to n. I have excluded the complements from the argument of the cost function for technical simplicity. In this situation, the worker's payoff is the sum of the wage received and the costs of congestion incurred. From the above, we have

$$E(u) = [a - b/(2m)] - c(n). \qquad (10)$$

Composing the solution for m from (8) into (10) gives

$$E(u) = [a - b/(2)] /[1/(F+a)) * [b/2 + [b^2/4 + bn(F+a)]^{1/2}] - c(n). \qquad (11)$$

(11) describes the familiar utility possibilities curve, giving the payoff to agents as functions of the size of city that they inhabit. The only wrinkle here is that the payoff is an expected utility because of the uncertainty inherent to matching. The assumption of convexity for c(-) and the boundedness of the first term imply that expected utility is a decreasing function of n for sufficiently large n. There is an agglomeration economy because the level of utility is initially increasing in n.

2.4. Interpretation

The most important aspect of this model's agglomeration economy is its generality. If we interpret the complements as employers, literally as entrepreneurs who arise from the population of workers, then we have a model where agglomeration economies stem from labor market pooling. Specialized workers are better matched with specialized entrepreneurs in thicker markets. If we instead interpret complements as the suppliers of specialized inputs, then we have a model where agglomeration economies arise from the sharing of these inputs. Or if we interpret the complements as specialized knowledge workers who transmit valuable knowledge to workers, we have a model where agglomeration economies arise from the exchange of knowledge. This is the place where the model's generality becomes most strained. The knowledge exchange captured here is not a Marshallian spillover in the sense that it is the "secrets of the trade" are not "in the air." There is no spillover because of the model's assumption of a bargain over the surplus created by the knowledge match between the two agents (see Callois (2008) for a recent model of industrial districts). This is not to say that knowledge exchange is always unmediated by markets. It sometimes is (see Helsley and Strange

(2004)). But the model does not capture spillovers in their pure form, which are certainly an important sort of agglomeration economies. It would not be difficult to re-specify the model to better capture spillovers. If we supposed, for instance, that each worker learns from the nearest complement and each complement learns from the nearest other complement (possibly probabilistically), this would again produce an increasing and concave benefit from agglomeration.

The model clearly establishes that there a many different logically consistent ways that one can model agglomeration. Agglomeration is thus more than just NEG. These many agglomeration economies have very similar empirical predictions, greater productivity, higher wages, spatial concentration, the creation of complements such as skilled labor, to name a few. The various microfoundations exhibit, in the words of Duranton and Puga (2004), "Marshallian equivalence." This means that it is likely to be much easier to persuasively document the existence of agglomeration economies than to conclusively identify their sources. Documenting existence is itself not easy, of course. Given the many possible sources of agglomeration, the scope for structural estimation seems to be limited to situations where one can be really certain of the sorts of agglomeration effects at work. Given the many other ways that one might generate an urban wage premium, for example, reduced form estimation requires attention be paid to endogeneity.

The modeling thus far has been phrased in terms of urbanization, the size of a city's population. It is straightforward to reinterpret the model in terms of localization, the degree of a particular industry's activity in a city. An immediate consequence of such a reinterpretation is that cities should specialize. If there are two industries – for example film and television production and automobile manufacturing – and these two industries correspond to separate characteristic spaces – with a given worker having an address on just one of them – then the benefits of agglomeration would depend only on the level of activity in a worker's own industry. There would be specialization. Alternatively, one could conceive of a different sort of specialization. Suppose that there are four industries, characterized by worker addresses in four quarters of the unit circle. In this case, with a sufficiently large aggregate population, there would again be specialization, with each specializing in one quarter of the unit circle. If there were not such a large aggregate population, however, it is possible that the equilibrium would involve incomplete specialization, with multiple industries co-agglomerating in a particular city. This is an issue that I will return to later in the paper.[2]

The analysis thus far has been static. The worker decision of whether or not to become active can instead be seen as describing dynamics. Adding worker dynamics to the model would then give a standard sort of growth model. The existence of workers leads to complements, which leads to more workers, and so on. A similar feedback mechanism is analyzed in Helsley and Strange (2001). The important point for use here is that growth dynamics can arise from any microfoundation, not just from knowledge spillovers as per Marshall (1890) or Jacobs (1969). It is frequently claimed that the observation of an urban growth relationship is evidence of knowledge spillovers. While such an observation is certainly consistent with knowledge spillovers, it is also consistent with other microfoundations such as labor market pooling or input sharing. This is a dynamic version of Marshallian equivalence.

The model, simple as it may be, sets out some predictions regarding the empirical implications of various sorts of agglomeration economies. The next section will discuss the empirical work in this area.

3. Estimating agglomeration effects

3.1. Background

Stepping back from the theoretical model, there are two fundamental questions regarding agglomeration economies. The first is: how strong are they (if they, in fact, exist at all)? The second is: what are their sources? In this section, I will discuss how economists have gone about answering these questions. As I noted above, I am not aiming here at an exhaustive methodological survey, but rather at identifying some successful recent approaches to identifying the existence and nature of agglomeration effects. I will primarily focus on newer research dealing with the labor market aspects of agglomeration and making use of data that are disaggregated by firm, worker, or geographical unit.

The theory gives us guidance as to where we might look for evidence of agglomeration effects. Last section's model has two especially sharp predictions. The first was that wage would be greater in thicker markets. This prediction has the fairly direct corollary that productivity and growth should also be associated with market thickness, The model's second prediction is that thick markets should display a rich range of complementary activities relative to thinner markets (i.e., more entrepreneurial activity or specialized input supply). The model does not itself say whether the thickness that matters should be associated with the size of a city or with the concentration of an industry. This distinction may not be a sharp one: industries should coagglomerate when they can draw on similar complementary activities. With regard to both predictions, the model does not say whether "market" should in this analysis be equated with a city (metropolitan area). That is certainly the tendency among researchers in urban economics, but it is not the only conceivable scale at which agglomeration effects might operate.

3.2. Productivity

In the model, wage is positively associated with agglomeration because productivity is. It is therefore natural to begin by looking directly at productivity. This approach has been taken by Moomaw (1981), Nakamura (1985), Henderson (1986) and others. Their models involved adding shifters to production functions that depended on the scale of activity within a city. The scale variables included the city's total level of employment and the level of employment in particular industries. The latter captured the urbanization economies that can explain the formation of large cities, while the former captured the localization economies that can explain the formation of industry clusters. The key finding of these early papers was there existed a relationship between both urbanization and localization and productivity, with the relationship differing across industries. More recent estimates of production functions can be found in Henderson (2003), while Graham (2007) carries out the parallel exercise of estimating the impact of agglomeration on congestion costs, a negative element in net productivity.

There are important estimation issues that arise in attempting to identify agglomeration economies. Combes et al (2008a) refer to two sorts of endogeneity problem. The first of these is an endogenous "quality of labor," with cities an industry clusters attracting workers with high levels of unobserved skills. The second is an endogenous "quantity of labor," with high productivity locations attracting workers and firms. Both sorts of endogeneity will lead to ordinary least squares (OLS) estimates of agglomeration effects that are biased upwards. Combes et al (2008a) employ geological instruments for the quantity of labor, obtaining estimates of agglomeration effects on total factor productivity (TFP) that are only slightly reduced.[3] Using finely disaggregated French panel data, they find a much more important effect due to endogenous quality. In all specifications, agglomeration effects continue to be present.

Dealing with endogeneity problems through instrumental variable and differencing methods is, of course, standard in applied economics. The nature of equilibrium in a system of cities makes the endogeneity problem especially daunting. Workers and firms are both mobile, as are the goods that they produce. The processes in question operate at a fairly local level -- this will be discussed further below -- so there are significant challenges to finding plausible instruments and disaggregated data that allows differencing. This means that in agglomeration research continues to involve attempts to find better instruments or better disaggregated data. A notably different approach is taken by Greenstone et al (2008). They use data on the locations of "million dollar plants" (MDP) that describes the locations that were not chosen. Prior to the MDP's arrival, the TFP of incumbents in these places followed the same trend as in the locations that were chosen. Afterwards, the TFP of plants in chosen locations rose, evidence of an agglomeration effect.

In any case, direct measures of production functions are often – perhaps even usually – impossible due to limitations in data. This has led to empirical work that looked for indirect evidence of agglomeration forces. This involves looking at two sorts of duals to the production function: growth and factor prices. Glaeser et al (1992) find a strong relationship between urban diversity and growth in estimates that aggregate across industries. Henderson et al (1995) estimate separately for individual industries. Some industries exhibit a positive relationship between diversity and growth. A set of mature industries have instead a relationship between industrial concentration and growth (localization). Combes et al (2004) consider growth using very rich French panel data. Dynamic panel methods allow them to identify separately area-industry fixed effects and static and dynamic externalities. They find consistent evidence of static externalities, but not dynamic externalities. The effect of local industrial composition is nuanced: there are spillovers across industries, so diversity matters, but an economy made up of several large industries rather than many small ones appears to be most conducive to growth. The impact of various agglomeration factors on urban growth in Brazil has been considered more recently by da Mata et al (2007).

Factor prices are impacted by agglomeration as well. In Section II, productivity increases the wage of both workers and complements. There was no land market in the model, but an increase in productivity would have also increased land prices (Roback (1982)). Rent data are generally very poor in a number of important dimensions: unobserved characteristics, excessive aggregation, market thinness. This has led researchers to employ the still imperfect but better data on wages.[4]

3.3. The urban wage premium

A long list of relatively recent papers have focused on urban wages. The list includes Glaeser and Mare (2001), Wheaton and Lewis (2002), Wheeler (2006), Lee (2007) and many others. There are several especially salient findings in this literature. First, both the level of overall activity is strongly related to agglomeration.   The striking fact presented by Glaeser and Mare is that workers in cities larger than 500,000 have wages that are 33 percent higher than workers outside of cities entirely.   Second, this urban wage premium shrinks significantly when one controls for worker characteristics.   Glaeser and Mare address unobserved heterogeneity in worker types in several ways. Arguably, the most complete is the estimation of a model with worker fixed effects using the NLSY. In this model, the wage premium remains significant, but shrinks to between 5 – 11 percent (for the PSID and NLSY models, respectively).

There has been less work (surprisingly) on the relationship between the concentration of individual activities and wage. Wheaton and Lewis (2002) show a positive relationship between wage and the local share of national employment in the worker's industry and also the share of national employment in the worker's occupation. These results are the equivalents of Glaeser and Mare's raw urban wage premium calculation. Wheaton and Lewis argue that there is no reason for a worker's unobserved abilities to be correlated with their localization measures. This claim can be assessed directly by considering both localization and urbanization.

This is exactly what Combes et al (2008) do. They employ a panel of French data in a way that essentially allows them to consider the full range of issues that had been considered separately in prior work. In France as in the U.S., there is a substantial premium associated with the largest cities. The average wage in Paris is roughly 15% higher than in other large cities such as Lyon. It is roughly 60% larger than in non-urban locations. As noted above, this finding is consistent with the existence of an externality, the selection of high-skill workers into cities, or both. It is also consistent with idiosyncratic advantages for particular locations.

In order to identify the three effects, Combes et al employ a two step procedure using disaggregate panel data. In the first stage, they estimate individual worker wage models, with various location and worker fixed effects. The worker fixed effects prove to be of great importance, but the place effects (both externalities and location advantages) are also strongly related to the wage. The second stage involves evaluating the correlates of the location fixed effects. Measures of urbanization are shown to be strongly related, while measures of localization are statistically significant but of economically modest magnitude. The pure place effects are similar. In net, we have strong evidence of endogenous quality of labor, or that "sorting matters", but also strong evidence for a relationship between urbanization a wage correcting for this. Fu and Ross (2007) also consider sorting and the urban wage premium. Their approach is to use place of residence to instrument for unobserved worker characteristics. As with Combes et al, they continue to find a premium. [5]

3.4. The geography of the urban wage premium

Combes et al treat cities as the markets within which agglomeration effects are present. The cities are defined for as employment area (*zones d'emploi*). There are 341 of these across France, typically defined as metropolitan areas. This approach is common to nearly all work on urban wages and, in fact, to most empirical work on agglomeration. This is sensible, since working at the level of the labor market fits theories of agglomeration like the matching model from Section II. This sort of exercise is also a necessary one to understand the agglomeration of economic activity into cities. In any case, this is an inescapable accommodation, since data are available at this degree of refinement.

As sensible as it is, examining agglomeration at the city level fails to allow for obvious agglomerative processes within cities. As noted in the introduction, there is great variation in density between city centers and suburbs. This is physically manifested in skyscrapers downtown, tall buildings in midtown and in important sub-centers, and much smaller buildings elsewhere in the city. See Redfearn (2007) or Berliant and Wang (2008) on subcenters and Helsley and Strange (2007) on skyscrapers. To really understand agglomeration, it is essential to understand geographic scope of agglomeration economies. This requires data that are spatially disaggregated.

Rosenthal and Strange (2003, 2005) consider urban entrepreneurship at the within-city level. The former considers the entire U.S. for six select industries. The latter considers the New York metropolitan area for a broader set of industries, including various service industries. The analysis involves estimating a kind of local growth model, specifically one that considers the determinants of a zipcode's rate of birth for new establishments. The models can be seen as models of complementary activity, specifically entrepreneurship. The two papers find evidence of industrial localization and also of the importance of local diversity. The most important result is that the effects are spatially localized. The effect of additional activity beyond five miles is on the order of half as strong as the effect inside.

Rosenthal and Strange (2008) take geographically flexible approach to the estimation of urban wage models. The flexibility is allowing for agglomeration effects to operate at the sub- or supra-city level. More specifically, the paper employs geographic information systems (GIS) to process the data in a way that allows for the characterization of a local economic environment within five miles, between five and 15 miles, and for various distances beyond that. This sort of flexibility is possible only with geographically disaggregated data. The paper's key result is the estimation of attenuating agglomeration economies. Endogeneity issues are dealt with using geological instruments and differencing. Regarding urban wages, the paper estimates urban wage premiums for activity within five miles of roughly 4 percent in a model with controls for worker abilities and with MSA fixed effects. Nearby activity matters, even controlling for effects that operate at the city level.[6]

The paper also considers human capital externalities, spillovers generated by educated workers. In considering this issue, the paper builds on prior work by Rauch (1993), Moretti (2004), Ciccone and Peri (2006), Fu (2007), Liu (2007), and others. The paper finds robust evidence of attenuating human capital effects. Transforming 100,000 less-than-college workers within 5 miles into college-educated – equivalent to the 25/90 difference in percentile – would increase the wage of a typical worker by 12 percent in an OLS model and by 15 to 30 percent in a range of instrumental variable models. These effects have a still larger impact on individuals who themselves have a college degree. If the transformation took place 5 to 25 miles outside of the individual's workplace, the estimated effects would be only half as large. Thus, for human capital effect, proximity matters. Similarly, Baldwin et al (2008) find spatially bounded agglomeration effects in Canadian data, and Gibbons and Silva (2008) show that pupil attainment is related to urban density.

3.5. Skills

The observation of a connection between skills and agglomeration goes back to Marshall (1890), at least. Econometric treatments of skills have typically equated a worker's skills with the level of education. Clearly, there is a relationship between skills and education. Equally clearly to any professor, the relationship is a complicated one. Some graduates exhibit considerable skill. Others exhibit not so much. In any case, education is a vertical measure, while the concept of skill has both vertical and horizontal dimensions.

Bacolod et al (2008a, 2008b) employ the Dictionary of Occupational Titles (DOT) to consider both the horizontal and vertical dimensions of skills. The DOT is designed for job-seekers, characterizing the skill requirements of a job. A professor would require both cognitive and social skills, but not motor or physical strength. A doctor would require cognitive, social, and motor skills, but not strength. And so on. Under the conditions of a hedonic equilibrium in a labor market, a worker is matched to a job requiring exactly his/her skills. This is a further sort of disaggregation, in this case down from the aggregate category of skilled workers.

Bacolod et al (2008a) use this hedonic attribution of skills to workers to consider which sorts of skills have prices that are positively related to city size. The key result of the paper is that urban wage premium is not enjoyed by all workers equally. The prices of cognitive and social skills increase with city size. The prices of motor skills and physical strength do not. This result is robust to a series of corrections for unobserved heterogeneity in worker quality. These include controlling for intelligence and social adjustment using the AFQT test and Rotter Index, controlling for quality of education by using measures of university selectivity, and estimating fixed effect models.

The results of Bacolod et al (2008a) on the distribution of skills are somewhat surprising. As per Section II's model, there are more high and low skill workers in larger cities (complements). Average skills levels are higher in larger cities, but not to an especially great extent. For instance, looking at mathematical skills (*gedm* in the DOT), cities of over 4 million have 1 percentage point more of their populations with algebra or more than the small cities with populations between 100,000 and 500,000. The difference in college education rates is 6 percentage points. This pattern repeats itself for other skills: a positive but not especially strong relationship between city size and skill levels. One possible explanation of this pattern is that there are important complementarities in production between the most and least skilled workers. This tempers any tendency for the highly skilled to agglomerate.[7]

Bacolod et al (2008b) focus on soft skills. The paper employs several of the non-cognitive measures from the DOT to consider the relationship between various sorts of people skills and agglomeration. These skill measures relate to the worker's ability to interact, either as a manager or outside of an authority relationship. This sort of interaction has long been considered to be a fundamental aspect of agglomeration economies. The idea that cities are about interactions is central to Jacobs (1969) and Vernon (1960) non-mathematical work, and it is also at the heart of research on spatial interactions, as exemplified by Ogawa and Fujita (1980) and Fujita and Ogawa (1982). It is tempting to believe that cities and industry must be populated with the most interactive of the economy's workers.

Bacolod et al (2008b) paint a more nuanced picture. Within-industry average levels of soft skills are indeed shown to rise with city size, although the effect is of modest magnitude. Average soft skill levels do not, however, rise with the city's share of national employment in a given industry (i.e., localization). In addition, the workers at the top of the skill distribution in large cities typically have higher levels of soft skills than in small cities, but the least skilled workers are less skilled in large cities than in small cities. This suggests that cities are about interaction, but only for the most skilled urban workers. Industry clusters do not attract especially interactive workers. One possible explanation for this is that agglomeration allows for two different sorts of increasing returns: interactions such as knowledge spillovers (the actual matches that take place in Section II's model) and a highly refined division of labor (the distance in characteristic space across which matching takes place). Soft skills enhance the former, but the specialization of a thick market makes the latter less necessary.

3.6. Other aspects of urban labor markets.

In discussing the empirics of agglomeration, I have thus far considered two aspects of agglomeration: the relationship to wage (and so productivity) and the relationship to various complementary activities. If one is willing to interpret the model more broadly, there are other implications that one can see for Section II's model on labor markets.

The first of these is turnover. As with all matching models, my model can be seen as relating to marriages. Since the agents are presented as being unattached prior to matching, the marriages are essentially first marriages. Turnover is about second marriages. In this situation, we would have agents moving from one match partner to another. Formally, one could conceive of an exogenous rotation of a worker's address in the characteristic space so that the worker may afterward be in a position where another match partner would create greater value. Even a small rotation may result in considerable turnover in a thick market, hence the positive relationship between turnover and agglomeration.

This sort of turnover (albeit, not in a matching context) is considered by Fallick et al (2006). They consider the much noted "job-hopping" that characterizes the Silicon Valley. They do find that college-educated computer workers in the Silicon Valley change jobs frequently relative to other similarly educated computer workers. Some of this difference can be attributed to a California fixed effect. Fallick et al argue persuasively that this effect is related to the limitations in enforceability of non-compete agreements in California. These agreements restrict employees from working in the same industry and location when they move, thus serious limiting mobility. Interestingly, they do not find a parallel California effect in other industries. This may reflect the innovativeness of computers and the consequent benefits for job-hopping. This illustrates a crucial and often underappreciated point: different industries agglomerate for different reasons. It is important to respect this heterogeneity, both as an economist and as a policy maker. See Bleakley and Lin (2007) and Freedman (2008) for more recent work on job hopping and agglomeration.

Labor supply is also related to agglomeration. In the model, workers supply a fixed amount of labor. That need not be the case. A frequently observed characteristic of large cities is that they are places of great energy, of striving and ambition, and of hard work. Rosenthal and Strange (2008) observe that there are several reasons for this. Since agglomeration raises wage, workers may be disposed to work harder. Or the workers who choose cities may be inherently disposed to hard work. Finally, the economic environment in cities may be marked by greater rivalry, a kind of urban rat race.

Rosenthal and Strange (2008) consider empirical evidence on the relationship between agglomeration and labor supply. The simple descriptive statistics are that hours worked are greater in large cities and also in locations where an industry is concentrated. These effects are not, however, constant across the labor force. The agglomeration effect is present for professional workers but not for nonprofessionals. Furthermore, the effect is much stronger for young professionals (30-40 years old) than for older professionals. It is stronger still when there are benefits to advancement in the local labor market, as captured by measures of the steepness of the within-occupation wage profile. While it is certainly true that both agglomeration-productivity effects and selection are at work, it is difficult to see this pattern as obtaining without some sort of rat race effect.

A final noteworthy aspect of the relationship between agglomeration and urban labor markets concerns monopsony. Manning (2007) shows that contrary to some theory, plant size is greater in denser markets. He shows that this can be explained by monopsony, where firms in cities have less power in the labor market and so tend to be larger. Wheeler (2006) offers an alternate explanation that stresses labor market pooling, specifically that large plants can take better advantage of flexible local markets.

3.7. Microfoundations

This section began with two questions. The literature reviewed so far has addressed the first – are there agglomeration economies? – and not the second – what are the microfoundations of agglomeration economies? To consider the sources of urban increasing returns, requires different methods. There are two broad ways that one might approach the issue. The first would be to examine one particular agglomeration economy such as input sharing, labor market pooling, or knowledge spillovers. An ideal empirical analysis of microfoundations requires highly disaggregated data. A direct estimate of the benefits of labor market pooling, for instance, would require data on matches between worker and firm, both actual and potential. In addition, all the requirements of estimating production functions (i.e., input data) would continue to be in force. The second approach would be to look simultaneously at several possible agglomerative forces in hopes of identifying their relative strengths. In either approach, it is necessary that the existence of a particular microfoundation must make a testable prediction that differs from other microfoundations. Neither approach is free from the endogeneity issues discussed above.

There has been considerable research on various individual microfoundations. Holmes (1999) finds that purchased inputs are a greater fraction of value where industries are concentrated, a finding consistent with input sharing. As he notes explicitly, it is possible that causality might instead or in addition run from urbanization to input sharing. Jaffee et al (1993) show that patent citations are spatially concentrated, a result consistent with local knowledge spillovers. More recent contributions on the relationship between agglomeration and patenting include Carlino et al (2007) and Agarwal et al (2008). Lin (2007) shows spatial concentration of "new work," a term coined by Jacobs (1969), by considering revisions of DOT (i.e., the emergence of new occupations). Charlot and Duranton (2004) show that being in a larger and more educated city is positively associated with workplace communication. Regarding labor market pooling, Simon (1988) shows that unemployment is greater in a specialized city, while Diamond and Simon (1990) show that wages are higher the greater is the cyclical variability of an industry's employment. The Costa and Kahn (2001) result that "power couples" (both spouses college educated) are increasingly located in large cities is also consistent with labor market pooling. More recently, Andersson et al (2007) identify benefits consistent with matching.

An alternate approach is to run some sort of "horse race" that considers the relative importance of various Marshallian microfoundations. Audretsch and Feldman (1996) and Rosenthal and Strange (2001) do this by regressing measures of agglomeration on industry characteristics that proxy for various sorts of microfoundations. The Rosenthal and Strange analysis is carried out at the zipcode, county, and state levels. The proxies include prior innovation in the industry (knowledge spillovers), manufactured and service input intensities (input sharing), managers per production worker and education levels (labor market pooling, knowledge spillovers), and the perishability of output (a proxy for transportation costs). The pattern of results is that shipping-oriented inputs (manufactured inputs, resources, perishability) influence agglomeration at the state level, knowledge spillovers impact highly localized agglomeration, and labor factors impact agglomeration at all levels of geography.

An alternate approach -- one that is directly implied by Section II's model -- is to use the pattern of co-agglomeration by industry to assess the importance of various microfoundations. The idea is that if one observes industries drawing on a certain complementary activity have a tendency to choose the same city, then one can argue that the complementary activity is generating agglomeration economies. Ellison et al (2007) look at several sorts of complementary activity. These include the industrial mix from the input-output table (input sharing), the labor mix by occupation (labor market pooling), a matrix of technology flows (knowledge spillovers). The find evidence of all three effects, with the strongest evidence for input sharing and then labor market pooling.

In a similar vein, in their work on million dollar plants, in addition to looking for externalities, Greenstone et al (2008) also look for evidence of their sources. They find that MDPs have larger effects on incumbent plant TFP when the incumbent plant employs similar workers or draws on a similar knowledge base. They do not find evidence consistent with input sharing. This exercise has estimated the effects of large plants on the local environment, so the results apply to a subset of all agglomeration economies. Having said that, the natural experiment method addresses possible reverse causality issues, a very important advance.

A related approach to determining the forces most responsible for agglomeration is simply to ask. Strange, Hejazi, and Tang (2006) effectively do this by employing survey data from Statistics Canada. The survey asks, among other things, for firms to identify key local competitive factors. The list of potential factors relates to various theories of the microfoundations of agglomeration economies. The key finding is that firms for which skills and innovation are deemed to be important are more likely to agglomerate. So are firms that face uncertain environment. This microfoundation was proposed explicitly by Vernon (1960), reappearing in Jacobs (1969). The idea is that cities are beneficial when production involves adaptation, a point obviously related to Section II's model.

Most of these approaches look for broad patterns of microfoundations across all industries. Given the many possible sources of agglomeration economies and the many obvious differences between industries, there is a strong case for looking for evidence of agglomeration effects within industries, yet another sort of disaggregation. This has the additional advantage of allowing for controls that capture especially important aspects of the industry in question. Sorenson and Audia (2001) and Klepper (2007) employ data on the evolution of, respectively, the shoemaking and the automobile manufacturing industries. They both find spinoffs from existing producers to be the crucial dynamic in the growth of local industry clusters. Holmes (2005) establishes the importance of supply chain effects in a study of sales offices. Garicano and Hubbard (2007) show a positive relationship between law firm specialization and the scale of the local market.

It is clear from the empirical work discussed in this section that there is a large and growing body of evidence that agglomeration economies exist. This is not to dismiss either natural advantage or a selection-productivity relationship, which have also been shown to lead to agglomeration. There is evidence of a range of agglomerative forces being at work, including the elements of NEG models, input sharing, knowledge spillovers, labor market pooling, and the management of risk. There is no single explanation for urbanization, and there is certainly no single explanation that explains the clustering of all industries.

4. Urban policy

It is commonly written that cities and industry clusters are important as engines of productivity growth. It is also commonly written that the importance of cities and clusters implies a dire need for some sort of policy intervention. This paper's review of research on agglomeration economies is consistent with the first view. There are many forces that can explain a relationship between agglomeration and growth, and there is a large body of empirical work that directly establishes a direct relationship. All of this points to the existence of a spatial division of labor, where the productivity of the macro-economy depends on the allocation of tasks between large cities and less dense locations and between industry clusters and less specialized locations.

The research reviewed here does not, however, offer especially strong support the second view, that policy intervention is warranted. It order to argue for such intervention it is necessary to identify a failure of the market allocation of resources relative to what the public sector might accomplish. To consider this possibility, I will return to the model and ask two questions. Given interactions as modeled in Section II, is the allocation of an aggregate population between cities efficient? Given a local population, is the interaction between agents efficient? I will deal with these issues briefly here because Duranton (2008) has discussed them much more completely, albeit in a different model of agglomeration.

Beginning with the first question, suppose that there exists an aggregate population N to be divided into cities. As long as the congestion costs associated with city population are convex, the model from Section II generates an inverted U-shaped utility possibilities curve. So would any other reasonable model of the microfoundations of agglomeration economies. If we suppose that N is large enough that we can ignore integer problems, an optimal city will be of size $n^* = \arg\max E(u)$, the maximum of the utility possibilities curve. It is easy to see that individual migration decisions need not lead to optimally sized cities. If a system of cities were made up of cities all smaller than n^*, individual migration would move towards the optimum.[8] However, if the system were made up of cities greater than n^*, there is no such pressure. If the "autarky" utility level outside of cities is u^0, the it is possible to have the equilibrium system of cities be comprised of cities with populations n^0 giving $E(u) = u^0$. In this equilibrium, cities are inefficiently large. More generally, it is possible to sustain cities for any population level in the interval [n^*,n^0]. This is a common finding in models of systems of cities. See, for instance, Henderson and Venables (2008).

Turning now to the second question, it is easy to see that the tendency for equilibrium cities to be too large is not the only inefficiency present in the model. In deciding whether to become active, a complement ignores the effect on the expected utility of workers and considers only his or her own payoffs. Equilibrium complement activity is thus inefficiently low for any population of workers. In other words, when one considers agglomerations as arising from some sort of interaction, then there is too little interaction in equilibrium. *Mutatis mutandis* this means that the first best optimum population of workers would be larger than n^*.

For all these reasons, the existence of agglomeration economies is a market failure in the classic sense. It is thus inappropriate to appeal to the welfare theorems and end the discussion of urban policy. One must acknowledge that there is at least some potential for a Pareto improvement. Rather surprisingly given the complexity of agglomeration effects, such a Pareto improvement appears to be exactly what Greenstone and Moretti (2004) appear to have found. Using the same data as in Greenstone et al (2008), they find that the jurisdictions that attract "million dollar plants" experience a positive break in property values of slightly more than 1%. Since property values incorporate both the external benefits and the locally borne costs of bidding for such plants, this is consistent with a local welfare increase. Of course, this is only one sort of local development program, and it is unclear whether there are significant costs of local tax competition that are borne by higher levels of government.

The story of Frederick Terman is instructive (Leslie and Kargon (1996)). He served as Dean of Engineering at Stanford, among other roles. William Hewitt and David Packard were some of his more notable his students. He is widely credited with playing an instrumental role in the creation of the Silicon Valley by encouraging knowledge spillovers and more generally the deployment of Stanford's human capital in high-technology entrepreneurship. Later in his life, he led an initiative to create another high-technology cluster in New Jersey. The circumstances were certainly propitious, with New Jersey sharing the Silicon Valley's endowments of an educated workforces, a strong university, access to important capital markets, and existing high-technology activity. And New Jersey had one advantage: experience at building a high-technology cluster. Unfortunately, the initiative was unsuccessful. New Jersey did continue to develop in high-technology sectors, but it did not follow anything like the Silicon Valley's explosive trajectory. This tale should be taken as highly cautionary by any policymaker who believes it will be easy to harness agglomeration economies and create an industry cluster.

In general, the critical problem facing the designers of urban policy is deciding on exactly the form that intervention should take. Unfortunately, the key conclusion from the empirical section is that the exact nature of agglomeration economies remains unclear. I am willing to accept as the state-of-knowledge that there exist external economies associated with city size. There is evidence arguing that the strongest effects are within industries. There is other evidence arguing that the strongest effects are associated with urbanization, either the scale of overall activity or its diversity. In other words, the old localization vs. urbanization debate is not completely settled. Neither is the equally old question of which forces explain the patterns of agglomeration. As a body of work, it seems reasonable to believe that there exists evidence for the existence of a range of agglomerative forces, both Marshallian and otherwise. It is also reasonable to believe that agglomerative forces differ in strength across activities. It is difficult to take this evidence, however, as providing clear guidance of how agglomeration economies should be managed. The appropriate management of knowledge spillovers is certainly different than is the management of labor market pooling. And input sharing is another process entirely.

This leaves me with only one policy prescription that I can endorse without qualification: first, do no harm (a Hippocratic Principle of Urban Policy). This does not argue at all for inactivity. Policies directed at public goods such as investing in education or infrastructure can be justified on many grounds. To the extent that the infrastructure and the educated workforce are urban, they will promote cities. These sorts of policies do not require the government to pick microfoundations or to pick "winning" activities for cities. The automobile industry is obviously one that involves strong agglomeration effects. It is also one that has been crucial to Canada's prosperity. The automobile industry has clearly benefitted from the provision of infrastructure that allows inputs to be shared locally, especially roads. Similarly, software has recently become a crucial industry, one that has benefitted from education policies. These policies are not limited in their impact. The same roads that allow input sharing in auto production also allow input sharing in other industries. And to the extent that highways are used for commuting, they also expand labor markets, thus encouraging agglomeration (Arnott (2007)). Similarly, the support of education is important well beyond software.

References

Agrawal, Ajay, Devesh Kapur, and John McHale (2008) 'How do spatial and social proximity influence knowledge flows? Evidence from Patent Data,' *Journal of Urban Economics*, 64, 264-269

Almazan, Andres, Adolfo De Motta, and Sheridan Titman, (2007) 'Firm location and the creation and utilization of human capital', *Review of Economic Studies* 74, 1305-1327

Andersson, Fredrik, Simon Burgess and Julia I. Lane (2007) 'Cities, matching and the productivity gains of agglomeration,' *Journal of Urban Economics* 61, 112-128

Arnott, Richard (1979) 'Optimal city size in a spatial economy,' *Journal of Urban Economics* 6, 65-89

Arnott, Richard (2007) 'Congestion tolling with agglomeration externalities,' *Journal of Urban Economics* 62, 187-203

Audretsch, David. B. and Maryann P. Feldman (1996) 'R&D spillovers and the geography of innovation and production,' *American Economic Review* 86, 630-640

Baum-Snow, Nathaniel (2007) 'Suburbanization and transportation in the monocentric model,' *Journal of Urban Economics* 62, 405-42.

Bacolod, Marigee, Bernardo Blum, and William C. Strange (2008a) 'Skills and the city,' Working paper

Bacolod, Marigee, Bernardo Blum, and William C. Strange (2008b) 'Urban interactions: Soft skills vs. specialization,' Working paper

Baldwin, John R., Desmond Beckstead, W. Mark Brown, and David L. Rigby (2008) ‘Agglomeration and the geography of localization economies in Canada,’ Working Paper.

Berliant, Marcus and Wang Ping (2008) ‘Urban growth and subcenter formation: A trolley ride from the Staples Center to Disneyland and the Rose Bowl,’ *Journal of Urban Economics* 63, 679-693

Bleakley, Hoyt and Jeffrey Lin (2007), 'Thick-market effects and churning in the labor market: evidence from U.S. cities,' Working paper

Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm (2007) ‘Looking for multiple equilibria when geography matters: German city growth and the WWII shock,’ *Journal of Urban Economics* 61, 152-169

Callois, Jean-Marc (2008) ‘The two sides of proximity in industrial clusters: The trade-off between process and product innovation,’ *Journal of Urban Economics* 63, 146-162

Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt (2007) ‘Urban density and the rate of invention,’ *Journal of Urban Economics* 61, 389-419

Cavailhès, Jean, Carl Gaigné, Takatoshi Tabuchi and Jacques-François Thisse (2007) ‘Trade and the structure of cities,’ *Journal of Urban Economics* 62, 383-404

Ciccone, Antonio and Giovanni Peri (2006) ‘Identifying human-capital externalities: Theory with applications,’ *Review of Economic Studies* 73, 381-412

Combes, Pierre-Philippe, and Henry G. Overman (2004) 'The spatial distribution of economic activities in the European Union,' *Handbook of Regional and Urban Economics*, Vol.4, 2845-2909, ed. J. V. Henderson and J. F. Thisse (Amsterdam: Elsevier)

Combes, Pierre-Philippe, Thierry Magnac, and Jean-Marc Robin (2004) 'The dynamics of local employment in France,' *Journal of Urban Economics* 56, 217-343

Combes, Pierre-Philippe, and Gilles Duranton (2006) 'Labor pooling, labor poaching, and spatial clustering,' *Regional Science and Urban Economics* 36, 1-28

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

Combes, Pierre-Philippe, and Gilles Duranton, Laurent Gobillon, and Sébastien Roux (2008b) 'Estimating agglomeration economies with history, geology, and worker effects,' Working paper

Córdoba, Juan-Carlos (2008) 'On the distribution of city sizes,' *Journal of Urban Economics* 63, 177-197

Costa, Dora L., and Matthew E. Kahn (2001) 'Power couples,' *Quarterly Journal of Economics* 116, 1287-1315

da Mata, Daniel, Uwe Deichmann, J. Vernon Henderson, Somik V. Lall, and Hyoung Gun Wang (2007) 'Determinants of city growth in Brazil,' *Journal of Urban Economics* 62), 252-272

Deng, Xiangzheng, Scott Rozelle, Jikun Huang and Emi Uchida (2006) 'Growth, population and industrialization, and urban land expansion of China,' *Journal of Urban Economics* 63, 96-115

Diamond, Charles A. and Curtis.J. Simon (1990) ‘Industrial Specialization and the Returns to Labor,’ *Journal of Labor Economics* 8, 175-201

Duranton, Gilles (2008) ‘From cities to productivity and growth in developing countries,’ *Canadian Journal of Economics*, forthcoming

Duranton, Gilles and Diego Puga (2001) ‘Nursery Cities: Urban Diversity, Process Innovation, and the Life-Cycle of Products,’ *American Economic Review*, 91, 1454-1477

Duranton, Gilles and Diego Puga (2005) ‘From sectoral to functional urban specialisation,’ *Journal of Urban Economics* 57, 343-370

Duranton, Gilles and Diego Puga (2004) ‘Micro-foundations of urban agglomeration economies,’ in *Handbook of Urban and Regional Economics,* Vol.4, 2063-2118, ed. J.V. Henderson and J.-F. Thisse, (Amsterdam: Elsevier)

Egger, Hartmut and Peter H. Egger (2007) ‘Outsourcing and trade in a spatial world,’ *Journal of Urban Economics* 62, 441-470

Elvery, Joel A. (2007) ‘City size and skill intensity,’ Working paper.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr (2007) ‘What causes industry agglomeration? Evidence from coagglomeration patterns,’ Working paper

Fallick, Bruce, Charles Fleischman and James B. Rebitzer (2006) ‘Job hopping in the Silicon Valley: The micro-foundations of an industry cluster,’ *Review of Economics and Statistics* 88, 472-481

Fu, Shihe (2007) ‘Smart café cities: Testing human capital externalities in the Boston metropolitan area,’ *Journal of Urban Economics* 61, 86-111

Fu, Shihe and Steven Ross (2007) ‘Wage premia in employment clusters: Agglomeration economies or worker heterogeneity,’ Working paper

Fujita, Masahisa (2000) ‘Thünen and the new economic geography,’ Working paper

Fujita, Masahisa, and Hideaki Ogawa (1982) ‘Multiple equilibria and structural transition of non-monocentric urban configurations,’ *Regional Science and Urban Economics* 12, 161-196

Fujita, Masahisa, and Jacques-Francois Thisse (2002) *The Economics of Agglomeration*, (Cambridge: Cambridge University Press)

Fujita, Masahisa, Tomoya Mori, J. Vernon Henderson, and Yoshitsugu Kanemoto (2004) ‘Spatial distribution of economic activities in Japan and China,’ in *Handbook of Regional and Urban Economics*, Vol .4, 2911-2977 ed. J. V. Henderson and J. F. Thisse (Amsterdam: Elsevier)

Garicano, Luis, and Thomas N. Hubbard (2007) ‘Specialization, Firms, and Markets: The Division of Labor Between and Within Law Firms’ Working paper

Gibbons, Stephen, and Olmo Silva, (2008*)* ‘Urban density and pupil attainment,’ *Journal of Urban Economics* 63, 631-650

Glaeser, Edward L. (1999) ‘Learning in cities,’ *Journal of Urban Economics* 46, 254-77

Glaeser, Edward L., and David C. Mare (2001) ‘Cities and skills,’ *Journal of Labor Economics* 19, 316-342

Glaeser, Edward L, Hedi Kallal, Jose A. Scheinkman, and Andrei Shleifer (1992), 'Growth in Cities,' *Journal of Political Economy* 100, 1126-1152

Graham, Daniel J. (2007) 'Variable returns to agglomeration and the effect of road traffic congestion,' *Journal of Urban Economics* 62, 103-120

Helsley, Robert W. (1990) 'Knowledge and production in the CBD,' *Journal of Urban Economics* 28, 391-403

Helsley, Robert W., and William C. Strange (1990) 'Agglomeration economies and matching in a system of cities,' *Regional Science and Urban Economics* 20, 189-212

Helsley, Robert W., and William C. Strange (1991) 'Agglomeration economies and urban capital markets', *Journal of Urban Economics* 29, 96-112

Helsley, Robert W., and William C. Strange (1994) 'City formation with commitment,' *Regional Science and Urban Economics* 24, 373-390.

Helsley, Robert W., and William C. Strange (1997) 'Limited developers,' *Canadian Journal of Economics* XXX, 329-348

Helsley, Robert W., and William C. Strange (2001) 'Innovation and input sharing,' *Journal of Urban Economics* 51, 25-45

Helsley, Robert W., and William C. Strange (2004) 'Knowledge barter in cities,' *Journal of Urban Economics* 56, 327-345

Helsley, Robert W., and William C. Strange (2007) ‘Agglomeration, opportunism, and the organization of production,’ *Journal of Urban Economics* 62, 55-75

Helsley, Robert W., and William C. Strange (2007) ‘A game-theoretic analysis of skyscrapers,’ *Journal of Urban Economics* 64, 49-64

Henderson, J. Vernon (1974) ‘The sizes and types of cities,’ *American Economic Review* 64, 640-656

Henderson, J. Vernon (1986) ‘Efficiency of resource usage and city size,’ *Journal of Urban Economics* 19, 47-70

Henderson, J. Vernon (2003) ‘Marshall's scale economies,’ *Journal of Urban Economics* 53, 1-28

Henderson, J. Vernon, and Yukako Ono (2008) ‘Where do manufacturing firms locate their headquarters?’ *Journal of Urban Economics* 63, 431-450

Henderson, J. Vernon, Ari Kuncoro, and Matthew Turner (1995) ‘Industrial development in cities,’ *Journal of Political Economy* 103, 1067-1085

Holmes, Thomas J. (1999) ‘Localization of industry and vertical disintegration,’ *Review of Economics and Statistics* 81, 314-25

Holmes, Thomas J. (2005) ‘The location of sales offices and the attraction of cities,’ *Journal of Political Economy* 113, 551-581

Holmes, Thomas J., and John J. Stevens (2004) ‘ Spatial distribution of economic activities in North America’ in *Handbook of Urban and Regional Economics*, Vol. 4. 2797-2843, ed. J.V. Henderson and J.-F. Thisse (Amsterdam: Elsevier)

Jacobs, Jane (1961) *The Death and Life of Great American Cities* (New York: Vintage)

Jacobs, Jane (1969) *The Economy of Cities* (New York: Vintage)

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca M. Henderson (1993) ‘Geographic localization of knowledge spillovers as evidenced by patent citations,’ *Quarterly Journal of Economics* 108, 577-598

Klepper, Steven (2007) ‘Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry,’ *Management Science* 53, 616-631

Krugman, Paul (1991) ‘Increasing returns and economic geography,’ *Journal of Political Economy* 99, 483-499

Lee, Sanghoon (2007) ‘Ability sorting and consumer city,’ Working paper

Lin, Jeffrey (2007) ‘Innovation, cities, and new work,’ Working paper

Liu, Zhiquing (2007) ‘The external returns to education: Evidence from Chinese cities,’ *Journal of Urban Economics* 61, 542-564

Marshall, Alfred (1890) *Principles of Economics*.(London: MacMillan)

Moomaw, Ronald L. (1981) ‘Productivity and city size: a critique of the evidence,’ *Quarterly Journal of Economics* 96, 675-688

Moretti, Enrico (2004) 'Human capital externalities in cities', in *Handbook of Urban and Regional Economics,* Vol.4, 2243-2292, ed. J.V. Henderson and J.-F. Thisse (Amsterdam: Elsevier)

Murata, Yasusada (2007) 'Taste heterogeneity and the scale of production: Fragmentation, unification, and segmentation,' *Journal of Urban Economics* 62, 135-160

Nakamura, Ryohei (1985) 'Agglomeration economies in urban manufacturing industries: a case of Japanese cities,' *Journal of Urban Economics* 17, 108-124.

Ottaviano, Gianmarco, Takatoshi Tabuchi, and Jacques-Francois Thisse (2002) 'Agglomeration and trade revisited,' *International Economic Review* 43, 409–435

Ogawa, Hideaki, and Masahisa Fujita (1980) 'Equilibrium land use patterns in a non-monocentric city,' *Journal of Regional Science* 20, 455-475

Pflüger, Michael, and Jens Sudekum (2008), 'Integration, agglomeration and welfare,' *Journal of Urban Economics* 63(2), 544-566

Quigley, John (1998) 'Urban diversity and economic growth,' *Journal of Economic Perspectives* 12, 127-138

Rappaport, Jordan (2008) 'A productivity model of city crowdedness,' *Journal of Urban Economics* 63(2), 715-722

Rauch, James (1993) 'Productivity gains from geographic concentration of human capital: Evidence from the cities,' *Journal of Urban Economics* 34, 380-400

Redfearn, Christian L. (2007) 'The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area,' *Journal of Urban Economics* 61(3), 519-54.

Roback, Jennifer (1982) 'Wages, rents, and the quality of life,' *Journal of Political Economy* 90, 1257-78

Robert-Nicoud, Frederic (2008) 'Offshoring of routine tasks and (de)industrialisation: Threat or opportunity—And for whom,' *Journal of Urban Economics* 63, 517-535

Rosenthal, Stuart S., and William C. Strange (2001 'The determinants of agglomeration,' *Journal of Urban Economics* 50, 191-229

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

Rosenthal, Stuart S., and William C. Strange (2004) 'Evidence on the nature and sources of agglomeration economies,' in *Handbook of Urban and Regional Economics*, Vol.4, 2119-2172, ed. J.V. Henderson and J.-F. Thisse (Amsterdam: Elsevier)

Rosenthal, Stuart S., and William C. Strange (2005) 'The geography of entrepreneurship in the New York Metropolitan Area,' *Economic Policy Review* 11, Federal Reserve Bank of New York, 29-54

Rosenthal, Stuart S., and William C. Strange (2008) 'Agglomeration and hours worked,' *Review of Economics and Statistics* 90, 105-115

Rosenthal, Stuart S., and William C. Strange (2008) 'The Attenuation of human capital spillovers,' *Journal of Urban Economics*, 64, 373-389.

Sato, Yasuhiro (2007) ‘Economic geography, fertility and migration,’ *Journal of Urban Economics* 61, 372-387

Simon, Curtis J. (1988) ‘Frictional unemployment and the role of industrial diversity,’ *Quarterly Journal of Economics* 103, 715-728

Smith, Adam (1776; 2008) *An Inquiry into the Nature and causes of the Wealth of Nations*, *(*Oxford: Oxford World's Classics)

Sorenson, Olav, and Pino G. Audia (2000) ‘The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States, 1940-1989,’ *American Journal of Sociology* 106, 424-62

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, 331-351

von Thünen, Johann (1863/1966) ‘*von Thünen's Isolated State: An English Edition of Der Isolierte Staat,’* (Oxford: Pergamon Press)

Vernon, Raymond (1960), *Metropolis 1985*, (Cambridge, MA: Harvard University Press)

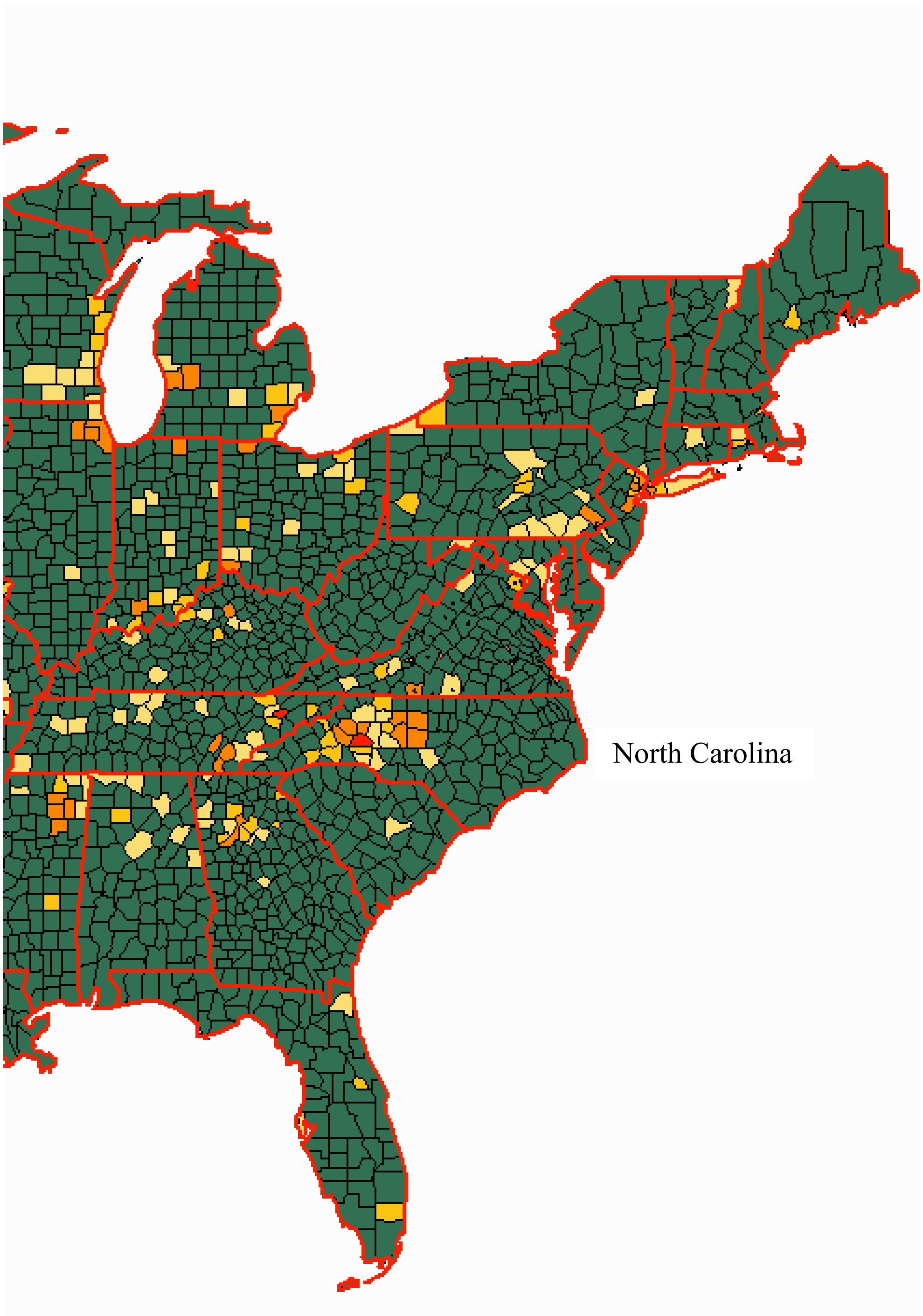
Wheaton, William C., and Mark J. Lewis (2002) ‘Urban wages and labor market agglomeration,’ *Journal of Urban Economics* 51, 542-562

Wheeler, Christopher H. (2006) ‘Cities and the growth of wages among young workers: Evidence from the NLSY,’ *Journal of Urban Economics* 60, 162-184

Zeng, Dao-Zhi (2008) ‘New economic geography with heterogeneous preferences: An explanation of segregation,’ *Journal of Urban Economics* 63, 306-324

Figure 1: Furniture Employment (SIC 25) Per Square Mile

Fourth Quarter 2002; Source: Dun and Bradstreet
Red: Greater than 10; Orange: 4 to 10; Dark Yellow: 2 to 3; Light Yellow: 1 to 2; Green: 0



Endnotes

Lead footnote: RioCan Real Estate Investment Trust Professor of Real Estate and Urban Economics. I gratefully acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada and the Marcel Desautels Centre for Integrative Thinking. I thank an anonymous referee for helpful comments. I also am grateful to the coauthors of my papers in this area, Robert Helsley, Stuart Rosenthal, Marigee Bacolod, Bernardo Blum, Greg DeCoster, Walid Hejazi, and Jianmin Tang

---

[1] Similarly, a specialized input supplier can arise from a population of producers.

[2] Instead of specialization in an industry, a city might specialize functionally. Duranton and Puga (2005) model this, and provide evidence of an increase in the tendency for functional rather than industrial specialization. Henderson and Ono (2008) examine the forces responsible for the agglomeration of headquarters.

[3] This follows Strange and Rosenthal (2008), a paper discussed later.

[4] An alternate approach is to look at the crowding that workers are willing to incur (i.e., Rappaport (2008). One could also look at the urban fooprint. See Deng et al (2008) for an example of this sort of analyis applied to a different topic.

[5]Lee(2005) considers the urban wage premium for doctors, an approach that deals with at least some of the unobserved heterogeneity problem since doctors performing similar tasks are present in cities of all sizes. His key result is that selection can explain the entire urban wage premium for this sector.

[6] All of these approaches look at geography in terms of distance. Baum-Snow (2007) shows that the infrastructure available in a city (specifically, highways) has important effects. His key finding is that the presence of additional highways is associated with less central city growth.

[7] See Elvery (2007) for a related vertical approach to skills making use of firm level data.

[8] Suppose all cities have size $n^0 < n^*$. Then an individual worker who moves to another city obtains a higher utility level. A more general version of this argument precludes the presence of any cities of smaller than optimum size in an equilibrium system. See Henderson (1974).