Coagglomeration and the Scale and Composition of Clusters

Robert W. Helsley*
Haas School of Business
University of California, Berkeley
Berkeley, CA 94720 USA

William C. Strange
Rotman School of Management
University of Toronto
Toronto, ON M5S 3E6 Canada

February 29, 2012

* We thank Hesham Abdel-Rahman, Alex Anas, Kristian Behrens, Marcus Berliant, Jan Brueckner, Edward Glaeser and Pratish Patel for helpful coments and suggestions. We also thank the participants of the 2011 Philadelphia Federal Reserve Bank – Princeton Conference on Urban and Regional Economics, the 2011 meetings of the Urban Economics Association, the 2011 ASSA meetings and seminar participants at HHI, LSE and UQAM for their useful comments. Finally, we thank the Social Sciences and Humanities Research Council of Canada and the Fisher Center for Real Estate and Urban Economics at UC Berkeley for financial support.

Abstract

Cities are neither completely specialized nor completely diverse. However, the theoretical agglomeration literature has focused almost entirely on the polar cases of complete diversity and specialization. In order to give theoretical foundations to cluster analysis, this paper develops a model that can also generate the intermediate case of cities that feature the coagglomeration of some but not all industries into clusters. The analysis sharply challenges the conventional wisdom that the size and composition of clusters are driven primarily by agglomerative efficiencies. It also illuminates what observed patterns of coagglomeration imply about the nature of agglomeration economies.

I. Introduction

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 than that of Los Angeles. The idea of business clusters, as popularized by Porter (1990), is that certain industries are related to each other, and these industries tend to locate together. Porter argues persuasively that this clustering is important for firm location decisions and for industrial policy. Ellison and Glaeser (1997) coined the term "coagglomeration" to refer to the more general tendency of various industries to locate together.

There is a growing empirical literature on the coagglomeration of industries into business clusters. Ellison and Glaeser (1997) document coagglomeration patterns, clearly establishing that there is both great variety in the forms that cities take and also important regularities in the equilibrium pattern of coagglomeration. Most importantly for our purposes, they show that cities are neither specialized one-industry towns nor homogeneous multi-industry economies. Ellison et al (2010) takes the additional step of using evidence of coagglomeration to provide evidence on the "relative importance" across industries of potential sources of agglomeration economies. This is some of the most powerful evidence in the literature on the microfoundations of agglomeration. All of this complements the vast literature on business clusters that began with Porter's work. Delgado et al (2010), Glaeser and Kerr (2010), Kerr (2010), and Klepper (2010) are recent empirical contributions.

There is only limited theoretical work on coagglomeration that would help us understand the scale and composition of business clusters. The theoretical literature on urbanization begins with Henderson's (1974) classic general equilibrium model. The heart of the paper is its characterization of the tradeoff between the benefits and costs of spatial concentration. This tradeoff is used to characterize optimal cities and relate their size to the nature of agglomeration economies. With urbanization economies, cities will be diverse. With localization economies, cities will be specialized. Regarding the intermediate case of coagglomeration, Henderson (1974, p. 641) writes that:

> [C]ities will probably specialize in bundles of goods, where, within each bundle, the goods are closely linked in production. They may use a common specialized labor force or a common intermediate input.

While Henderson (1974) has a careful analysis of the polar cases of specialization and diversity, he offers no formal analysis of coagglomeration.[1]

This paper provides such analysis, sharply challenging the conventional wisdom regarding coagglomeration and business clusters. As above, it is common to believe that the mix of activities in a cluster reflects the mutual benefits that arise from colocation. In contrast, this paper's theoretical analysis will show that equilibrium clusters are likely to be inefficient in both the scale of agglomeration and also in cluster composition.

The source of the key result is that individual migration is a weak instrument for promoting efficiency. To understand this heuristically, imagine a situation where a worker chooses a location taking other workers' location decisions as given. An efficient cluster balances the productivity benefits of agglomeration and coagglomeration against the costs of congestion. The worker may be willing to locate in an inefficient cluster if the alternative is forming a new city (autarky) and this offers lower utility. This is a coagglomeration corollary to the familiar migration pathology in the agglomeration literature (see Abdel-Rahman and Anas (2004)). In fact, the result is quite different in the coagglomeration case, with the Pareto efficient allocation no longer being consistent with equilibrium. As will be seen, this is a consequence of externalities between industries that arise in the coagglomeration case, but not in the traditional analysis.

The inefficiency is manifested in both the scale and the composition of business clusters. Even if industries are partitioned efficiently into clusters, the migration pathology will lead to excessive scale. Furthermore, migration incentives may fail to lead to an efficient partition. For instance, a partition made up of specialized cities can be sustained as an equilibrium as long as no worker is willing to be the first to deviate. Also, a cluster may include industries that do not benefit each other as long as it provides a high enough utility level to discourage any worker from deviating. This is quite possible, since workers may prefer to remain in a city with many workers in their own industry. In sum, individual migration fails to produce efficient cluster scale and composition.

---

[1]Neither does the vast literature on urbanization that follow Henderson. It is worth noting that the word "coagglomeration" does not appear in the indices of either the most recent *Handbook of Regional and Urban Economics* (Henderson and Thisse, eds. (2004)) or the comprehensive textbook on agglomeration by Fujita and Thisse (2002).

This inefficient coagglomeration result is potentially very important. The urban economics literature continues to embrace the idea that localization economies create industry towns, while urbanization economies create completely diverse cities. This claim is based on a strong informal welfare theorem that agglomeration should be efficient. By considering coagglomeration explicitly, we show otherwise. It is, unfortunately, quite possible to have equilibrium clusters that fail to involve efficient coagglomeration.

The result is also important because of its empirical implications. Audretsch and Feldman (1996) and Rosenthal and Strange (2001) estimate relationships between industry agglomeration and industry characteristics. The idea is that if agglomeration is based on, for instance, knowledge spillovers, then one should observe agglomeration of knowledge-intensive industries. In an important recent paper, Ellison et al (2010) employ the creative strategy of looking at the flows of ideas, people, and goods to better understand microfoundations. Our paper shows that it is possible that migration may fail to produce an equilibrium where industries locate together if and only if there are underlying mutual benefits. Potentially beneficial coagglomeration may not necessarily occur, and the coagglomeration that does occur may not be that which creates the greatest benefits. For example, industries that could benefit from each other's knowledge may not co-locate in equilibrium. This argues for caution in dismissing the relative importance of possible microfoundations based on either cross sectional patterns of industrial agglomeration or coagglomeration.

The remainder of the paper is organized as follows. Section II provides a more detailed review of the current state-of-knowledge regarding coagglomeration. Section III describes the basics of our formal analysis. Section IV analyzes the characteristics of an efficient cluster. Section V shows that an efficient cluster cannot be an equilibrium with free migration. Section VI considers the possibility of specialized cities. Section VII examines the matching of industries to cities. It thus considers the efficiency of cluster composition. Section VIII discusses the empirical implications of our results on coagglomeration. Section IX concludes.

II. Literature

The classics of the agglomeration literature all touch on coagglomeration only lightly. Marshall (1890) is, of course, primarily concerned with localization, the specialization of an area in one industry. There are only a few places where he explicitly considers coagglomeration. One is his analysis of the possibility that "supplementary" industries may co-locate:

> On the other hand a localized industry has some disadvantages as a market for labour if the work done in it is chiefly of one kind, such for instance as can be done only by strong men. In those iron districts in which there are no textile or other factories to give employment to women and children, wages are high and the cost of labour dear to the employer, while the average money earnings of each family are low. But the remedy for this evil is obvious, and is found in the growth in the same neighbourhood of industries of a supplementary character. Thus textile industries are constantly found congregated in the neighbourhood of mining and engineering industries, in some cases having been attracted by almost imperceptible steps; in others, as for instance at Barrow, having been started deliberately on a large scale in order to give variety of employment in a place where previously there had been but little demand for the work of women and children. (Marshall (1890), Book IV.X.10)

Marshall also identifies a statistical basis for efficient coagglomeration:

> A district which is dependent chiefly on one industry is liable to extreme depression, in case of a falling-off in the demand for its produce, or of a failure in the supply of the raw material which it uses. This evil again is in a great measure avoided by those large towns or large industrial districts in which several distinct industries are strongly developed. If one of them fails for a time, the others are likely to support it indirectly; and they enable local shopkeepers to continue their assistance to workpeople in it. (Marshall (1890), Book IV.X.10)

These passages, while without formal theory, are impressively precise in describing microfoundations for coagglomeration. They do not, however, have much to say about how a location equilibrium is reached or about coagglomeration as a general phenomenon.

Vernon (1960) and Jacobs (1969) are more interested in diverse cities, so their analysis has more to say about coagglomeration. In Vernon (1960), much attention is given to the nature of "increasing returns industries," those that can benefit from the size and diversity of New York's markets. In this analysis, co-location arises because of similarities between the industries (they need similar environments) rather than their differences (drawing on male vs. female workers, as in Marshall). In Jacobs (1969), "new work" arises from prior activities. For example, leather working arises from agriculture in early settlements, and sliding door manufacturing arises from airplane manufacturing in the modern era. In this case, there is little discussion of entrepreneurs choosing locations. Instead, co-location arises from the process of technological advance. Again, there is no general treatment of coagglomeration or of location equilibrium.

The situation is similar in research on business clusters. The literature on business clusters following Porter (1990) has paid more attention to coagglomeration, but without formal welfare analysis. Porter's early classics in this area have an explicitly normative dimension, focusing on clusters of related activities as sources of competitiveness that should be nurtured by the public sector. In a similar spirit, the dynamics of coagglomeration are central to Saxenian's (1994) account of how the Silicon Valley transitioned from microchips to software.

The theoretical literature on microfoundations has formalized these ideas, and other forces that might encourage coagglomeration as well. See Duranton and Puga (2004) for a survey. For the most part, this literature does not consider the forces that might lead to coagglomeration. When it does most of the analysis deals with the polar cases of specialization and diversity. Krugman's (1991a) analysis of labor market pooling is an interesting exception. It formalizes the Marshallian treatment of risk sharing. See also the extension in the Duranton and Puga survey. Another interesting exception is the Duranton and Puga (2001) analysis of "nursery cities," which provides a formal model of innovation in the spirit of Jacobs and Vernon. Its key implications for coagglomeration are that young firms and industries benefit more from diversity (coagglomeration with a lot of other industries) than do their more mature counterparts.

The theoretical literature that our paper draws from deals with systems-of-cities (see Abdel-Rahman and Anas (2004) for a survey). As noted above, Henderson (1974) is seminal. It presents analysis that can be taken as a model of the formation of a system of specialized cities or a system of diverse cities. Abdel-Rahman and Fujita (1993), Abdel-Rahman (1996), and Anas and Xiong (2005) consider the circumstances under which a specialized or a diverse system of cities will arise. These models have explicit microfoundations, but their specification means that their models do not consider intermediate coagglomeration, where a city produces more than one good but not all goods in the economy.[2] Abdel-Rahman and Fujita (1993) and Abdel-Rahman (1996) are primarily concerned with the issue of when equilibrium will be specialized or diverse rather than with efficiency. These two papers are written in the Henderson (1974) tradition. In contrast, Anas and Xiong (2005) builds on Anas and Xiong's (2003) departure from the

[2]Because of the monopolistic competition specification employed, models in the New Economic Geography tradition (i.e., Krugman (1991b)) typically consider cities that produce the entire range of goods.

Henderson tradition by explicitly incorporating trade into a model of urban development. This paper examines how differences in trade costs between intermediates and final goods can impact the sort of coagglomeration that occurs (i.e., horizontal or vertical).

Finally, it is worth saying a little more about what coagglomeration means. We have thus far employed standard usage by referring to the coagglomeration of industries, and we will continue to employ this usage below. However, the joint location of firms in different NAICS industries is certainly not the only pattern of coagglomeration that might arise. Duranton and Puga (2005) present a model showing that changes in technology can lead to an urban system based on functional specialization rather than sectoral or industry-based specialization. Various papers on urban labor markets have considered the possibilities of complementarities between workers with various levels of education (Ciccone and Peri (2006)) or horizontally differentiated types of skill (Bacolod et al (2008) and (2009)). All of this means that our paper's results can be interpreted as reflecting on various sorts of coagglomeration.

In sum, the literature has focused more on the microfoundations of coagglomeration than on the implications of individual migration for the efficiency of cluster scale and composition. The folk wisdom is that despite the various increasing return issues associated with city formation, coagglomeration is likely to be efficient. The rest of the paper will employ a system-of-cities model to consider this folk wisdom.

III. A model of clusters and coagglomeration

A. Elements

This section develops a model of the formation of business clusters. In these multi-industry cities migration depends on complementarities in production between workers in different industries. There are I worker types and industries, indexed by $i \in \{1,2,\ldots,I\}$. Only type-i workers are employed by firms in industry i. There are J cities. The number of type-i workers in city $j \in \{1,2,\ldots,J\}$ is denoted by $n_{ij} \geq 0$. The composition of local employment is given by the vector $(n_{1j},n_{2j},\ldots,n_{Ij})$. The population of city j is $N^j = \sum_i n_{ij}$. The aggregate number of type-i workers in the economy is $N_i = \sum_j n_{ij}$. We assume that the economy is large enough that integer issues can be ignored.

B. Firms

Firms and industries are perfectly competitive. Output of a firm in industry i in city j is

$$q_{ij} = g_i(n_{1j},n_{2j},\ldots, n_{Ij})m_{ij}, \tag{III.1}$$

where m_{ij} is the number of type-i workers employed by the firm. Output per worker $g_i(\cdot)$ is a continuously differentiable and strictly concave function of the composition of local employment. We assume that $g_i(\cdot)$ satisfies $\partial g_i/\partial n_{kj} > 0$, $\partial^2 g_i/\partial n_{kj}^{\ 2} < 0$, and $\partial^2 g_i/\partial n_{kj}\partial n_{k'j} = \partial^2 g_i/\partial n_{k'j}\partial n_{kj} > 0$ for all i, k, $k' \in \{1,2,\ldots,I\}$.[3] The key feature of this specification is that the marginal benefit of another worker of any type ($\partial g_i/\partial n_{kj}$) is increasing in the number of workers of other types in the local economy. In the language of supermodular games (Topkis (1998)), $\partial g_i/\partial n_{kj}$ has strict "increasing differences" in $n_{k'j}$ for $k' \neq k$, i, k, $k' \in \{1,2,\ldots,I\}$. The agglomeration economies that underlie $g_i(\cdot)$ are generally attributed to some combination of knowledge spillovers, input sharing and improved matching in thick input markets (Marshall (1890); see Duranton and Puga (2004) for a survey). Our model allows agglomeration economies to operate both within and across industries. In other words, the model allows for both localization effects, arising from the spatial concentration of workers in a particular industry, and urbanization effects, arising from the diversity of economic activity in the local economy as a whole.

In this setting, optimum and equilibrium development patterns will depend on the strength of complementarities across worker types for a given industry ($\partial g_i/\partial n_{ij}$ relative to $\partial g_i/\partial n_{kj}$, $k \neq i$) and across industries for a given worker type ($\partial g_i/\partial n_{ij}$ relative to $\partial g_k/\partial n_{ij}$, $k \neq i$). The empirical literature on agglomeration economies (see Glaeser and Gottlieb (2010) and Puga (2010) for recent reviews) does not allow precise comparisons of these effects. However, there is some evidence (e.g., Henderson (1986, 2003)) that the productivity benefits associated with spatial concentration are especially strong for own-industry employment. Thus, it seems most consistent with the evidence to suppose that the own-industry effects are stronger than the cross-industry effects in our model.[4]

---

[3] Strict concavity requires that the matrix of second partial derivatives of $g_i(\cdot)$ be negative definite. For I = 2, this requires $\partial^2 g_i/\partial n_{kj}^{\ 2} < 0$, k = 1,2, and $(\partial^2 g_i/\partial n_{1j}^{\ 2})(\partial^2 g_i/\partial n_{2j}^{\ 2}) - (\partial^2 g_i/\partial n_{1j}\partial n_{2j})^2 > 0$. Intuitively, the last condition says that the concavity in either type must be stronger than the complementarities between them.

[4] It is worth noting that this distinction is not the same as the traditional localization versus urbanization distinction. We are instead assuming that own-industry effects are large relative to effects associated with any particular other

Accordingly, we make two further assumptions about $g_i(\cdot)$:

Assumption 1: for $n_{ij} = n_{kj}$ for all $k \neq i$, $\partial g_i/\partial n_{ij} > \partial g_i/\partial n_{kj}$. (III.2)

This condition says that with equal numbers of all worker types, adding another type-i worker has a larger impact on productivity in industry i than adding another type-k worker, $k \neq i$.

Assumption 2: for any $(n_{1j},n_{2j},\ldots,n_{Ij}) > 0$, $\partial g_i/\partial n_{ij} > \partial g_k/\partial n_{ij}$, $k \neq i$. (III.3)

This condition says that for any employment vector $(n_{1j},n_{2j},\ldots,n_{Ij}) > 0$, adding another industry i worker has a larger impact on productivity in industry i than in industry k, $k \neq i$. It is important to note that even with these restrictions, our model allows the diversity of economic activity in the local economy to have a positive impact on productivity in any industry. We discuss the consequences of Assumptions 1 and 2 in detail below.

The profit of a firm in industry i in cluster j is

$$\pi_{ij} = g_i(n_{1j},n_{2j},\ldots,n_{Ij})m_{ij} - w_i m_{ij}, \qquad \text{(III.4)}$$

where w_i is the wage paid to a type-i worker. To parallel standard competitive models with single industry cities (see Abdel-Rahman and Anas (2004) for a review), we assume that all industries produce the numeraire commodity.[5] Profit maximization implies

$$w_i = g_i(n_{1j},n_{2j},\ldots,n_{Ij}), \qquad \text{(III.5)}$$

ensuring zero maximum profits for all firms. Since the wage function captures the key elements of the technology, we will not need to refer explicitly to the behavior of firms in the remainder of the paper.

---

industry. We are not making any assumptions regarding the sum of other industry effects, which is the traditional conception of urbanization economies.

[5] By treating output as numeraire, we are effectively assuming away trade. See Anas and Xiong (2003,2005) for analysis of the implications of trade for specialized and diverse cities. We discuss this issue in more detail in the Conclusion.

C. Workers

We assume that workers are perfectly mobile. Each worker chooses a city to maximize utility. The indirect utility of a type-i worker in city j depends on how the composition of employment -- its nature as a cluster -- impacts the worker's wage:

$$V_i(n_{1j},n_{2j},\ldots,n_{Ij}) = w_i - c(N^j) = g_i(n_{1j},n_{2j},\ldots,n_{Ij}) - c(N^j), \tag{III.6}$$

where $c(\cdot)$ represents the cost of living in a city with population N^j. We also allow workers to locate outside of a city and receive the autarky utility level, which we write as $V_i(0)$. The autarky utility level impacts the employment vectors that can be supported as Nash equilibria with free migration; this issue is discussed in detail below. We assume that the residential cost function $c(\cdot)$ is continuous, increasing and (at least weakly) convex.

It is conventional (e.g., Henderson and Becker (2000)) to suppose that $c(\cdot)$ equals expenditures on land (or housing) plus transportation in a simple model of residential location. For example, if workers derive utility from a numeraire good and land, and land consumption is fixed at one unit per worker, then spatial equilibrium (equal utility across locations) requires $r'(x) = -t$, where $r(x)$ is land rent, x is commuting distance and $t > 0$ is commuting cost per mile. If the opportunity cost of urban land is zero, then the land rent function can be written $r(x) = -t(x_b - x)$, where x_b is the boundary of the city. Finally, if the developed area is rectangular with one unit of land at each location, then the market clearing condition gives $x_b = N^j$, so,

$$c(N^j) = r(x) + tx = tN^j. \tag{III.7}$$

Other forms for the residential cost function are possible, depending on substitution in consumption, the geography, congestion, the redistribution of land rents, and other factors.

IV. Optimal clusters

A. Type optimal clusters

We use the term *type optimal cluster* to refer to the local employment vector that maximizes the utility of a worker of a particular type.[6] For example, the type-i optimal cluster $(n^{*i}_{1j}, n^{*i}_{2j}, \ldots, n^{*i}_{Ij})$ maximizes $V_i(n_{1j}, n_{2j}, \ldots, n_{Ij})$. From (III.6), the first-order necessary conditions that characterize a type-i optimal cluster are:

$$\partial V_i/\partial n_{kj} = \partial g_i/\partial n_{kj} - c'(N^j) = 0 \text{ for all } k \in \{1,2,\ldots,I\}.^7 \qquad \text{(IV.1)}$$

Thus, at a type-i optimum, the marginal benefit (to a type-i worker) of another worker of any type equals the marginal cost of accommodating population in the city. Since the residential cost function does not depend on type, this implies that, at the optimum, adding another worker of any type-must have the same, positive impact on the productivity of workers in industry i:

$$\partial g_i/\partial n_{ij} = \partial g_i/\partial n_{kj} \text{ for all } k \neq i. \qquad \text{(IV.2)}$$

Note that if the output per worker function $g_i(\cdot)$ is the same for all i, then the composition of local employment will be the same in all type optimal clusters. However, there is nothing in the empirical literature on agglomeration economies that suggests that own- and cross-industry effects are identical across industries. In the general case where the output per worker function $g_i(\cdot)$ differs by industry, the composition of employment in type optimal clusters will differ as well. For example, if own- and cross-industry effects are different for automotive repair than for financial services, then the composition of local employment that is optimal for a mechanic will be different from the composition that is optimal for an accountant, although both types of workers would presumably be present in both type optimal cities.

In fact, (III.2) and (IV.2) imply that a type optimal cluster cannot have equal numbers of workers of all types. From (III.2), for $n_{ij} = n_{kj}$ for all $k \neq i$, $(\partial g_i/\partial n_{ij})/(\partial g_i/\partial n_{kj}) > 1$. From (IV.2),

---

[6] Note that this approach relies on the assumption that the economy is large enough that integer issues can be ignored. Otherwise, there is a remainder of type-i workers who receive their autarky utility level, and a type-i optimal cluster would maximize $Jn_iV_i(n_{1j}, n_{2j}, \ldots, n_{Ij}) + (N_i - Jn_i)V_i(0)$, where the number of clusters J is the largest integer smaller than N_i/n_i.

[7] These conditions characterize an interior solution where $n_{ij} > 0$ for all i. If $g_i(\cdot)$ is strictly concave and $c(\cdot)$ is (at least weakly) convex, the necessary conditions (IV.1) are also sufficient.

$(\partial g_i/\partial n_{ij})/(\partial g_i/\partial n_{kj}) = 1$ for all $k \neq i$, at a type-i optimum. The total derivative of the ratio $(\partial g_i/\partial n_{ij})/(\partial g_i/\partial n_{kj})$ is

$$d[(\partial g_i/\partial n_{ij})/(\partial g_i/\partial n_{kj})] = [[(\partial g_i/\partial n_{kj})(\partial^2 g_i/\partial n_{ij}^2) - (\partial g_i/\partial n_{ij})(\partial^2 g_i/\partial n_{kj}\partial n_{ij})]/(\partial g_i/\partial n_{kj})^2]dn_{ij} + [[(\partial g_i/\partial n_{kj})(\partial^2 g_i/\partial n_{ij}\partial n_{kj}) - (\partial g_i/\partial n_{ij})(\partial^2 g_i/\partial n_{kj}^2)]/(\partial g_i/\partial n_{kj})^2]dn_{kj} \quad \text{(IV.3)}$$

Concavity, $\partial^2 g_i/\partial n_{kj}^2 < 0$ for all $i, k \in \{1,2,\ldots,I\}$, and complementarity, $\partial^2 g_i/\partial n_{ij}\partial n_{kj} > 0$ for all $i,k \in \{1,2,\ldots,I\}$, imply that the bracketed term multiplying dn_{ij} is negative, while the bracketed term multiplying dn_{kj} is positive. Thus, from a perfectly diversified composition where $n_{ij} = n_{kj}$ for all $k \neq i$, to move toward a type-i optimum (to decrease the ratio $(\partial g_i/\partial n_{ij})/(\partial g_i/\partial n_{kj})$) involves increasing n_{ij} or decreasing n_{kj} or both. This implies that industry i workers are in a plurality in a type-i optimal cluster: $n^{*i}_{ij} > n^{*i}_{kj}$ for all $k \neq i$. This analysis is summarized in the following proposition.

Proposition 1: Suppose, as in Assumption 1, that with equal numbers of workers of all types ($n_{ij} = n_{kj}$ for all $k \neq i$), adding another type-i worker has a larger impact on productivity in industry i than adding another type-k worker, $k \neq i$ ($\partial g_i/\partial n_{ij} > \partial g_i/\partial n_{kj}$). Then type-i workers are in a plurality in a type-i optimal cluster: $n^{*i}_{ij} > n^{*i}_{kj}$ for all $k \neq i$.

Clusters are assemblies of related industries. Proposition 1 means that the various inhabitants of a cluster have different preferences for the composition of local employment. Specifically, the nature of complementarities between industries as formalized in Assumption 1 implies that workers prefer that their own industry be dominant in the cluster. It is, of course, impossible for all workers to get what they want in this regard, and this will play an important role in the welfare economics that follow. This result is new to this paper; single-industry models of agglomeration cannot capture this sort of conflict in the composition of workers' ideal clusters.[8]

Type optimal clusters for I = 2 are illustrated in Figure 1, together with the elliptical level sets of the utility functions $V_i(n_{1j},n_{2j})$. For I = 2, Proposition 1 implies that type-i workers are in

[8]The Appendix presents closed form solutions for type optimal clusters in an example based on a specific form for $g_i(\cdot)$ from (III.1) and the linear form for $c(\cdot)$ from (III.7).

a majority in a type-i optimal cluster. Thus, in the figure, the type-1 and type-2 optimal cities lie below and above the 45° line, respectively.

B. Pareto efficient clusters

In this context, a *Pareto efficient cluster* is an employment vector $(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})$ such that it is not possible to increase the utility of a type-i worker without decreasing that of a type-k worker, for all i and for some $k \neq i$. It will be convenient to characterize such an allocation as a maximum of the weighted utility function[9]

$$W = \sum_i \lambda_i V_i(n_{1j},n_{2j},\ldots,n_{Ij}), \tag{IV.4}$$

where $\lambda_i > 0$ for all $i \in \{1,2,\ldots,I\}$ and $\sum_i \lambda_i = 1$. Letting $\lambda_I = 1 - \sum_{i\neq I} \lambda_i$, the first order conditions that characterize a Pareto efficient cluster are

$$\sum_i \lambda_i \partial V_i/\partial n_{kj} = \sum_i \lambda_i(\partial g_i/\partial n_{kj}) - c'(N^j) = 0 \text{ for all } k \in \{1,2,\ldots,I\}. \tag{IV.5}$$

$\partial g_i/\partial n_{kj}$ is the marginal impact of a type-k worker on the productivity of workers in industry i. $\sum_i(\partial g_i/\partial n_{kj})$ is thus the aggregate marginal impact of a type-k worker on the productivity of workers in all industries. (IV.5) states that in a Pareto efficient allocation, the weighted aggregate marginal benefit of a type-k worker equals the marginal cost of accommodating population in the city, for all $k \in \{1,2,\ldots,I\}$. The conditions in (IV.5) implicitly define a *contract surface* consisting of all Pareto efficient employment vectors for different weights $(\lambda_1,\lambda_2,\ldots,\lambda_I)$. Note that as λ_i approaches 1, and thus λ_k approaches 0 for all $k \neq i$, the conditions in (IV.5) approach (IV.1), the conditions that characterize a type-i optimal city. The contract surface thus connects or spans the type optimal cities in the economy.[10]

---

[9] To see that a maximum of W is Pareto efficient, suppose that $(n^*_1,n^*_2,\ldots,n^*_I)$ is a maximum of W, but is not Pareto efficient. Then there must exist some other employment vector $(n'_1, n'_2,\ldots,n'_I)$ such that $V_i(n'_1,n'_2,\ldots,n'_I) \geq V_i(n^*_1,n^*_2,\ldots,n^*_I)$ with strict inequality for some i. This implies $\sum_i \lambda_i V_i(n'_1, n'_2,\ldots,n'_I) > \sum_i \lambda_i V_i(n^*_1,n^*_2,\ldots,n^*_I)$, contradicting the assumption that $(n^*_1,n^*_2,\ldots,n^*_I)$ is a maximum of W. See Sundaram (1996), pp. 83. Note that for $\lambda_i = N_i/\sum_i N_i$ maximizing W is equivalent to maximizing aggregate welfare.

[10] A constrained Pareto efficient allocation is a point on the contract surface that also satisfies the aggregate population constraints. We characterize such an allocation in detail below.

A Pareto efficient cluster is easiest to visualize for the case of I = 2 industries. In this case, the first-order conditions become

$$\lambda\partial V_1/\partial n_{1j} + (1-\lambda)\partial V_2/\partial n_{1j} = \lambda\partial g_1/\partial n_{1j} + (1-\lambda)\partial g_2/\partial n_{1j} - c'(n_{1j} + n_{2j}) = 0, \qquad \text{(IV.6)}$$

$$\lambda\partial V_1/\partial n_{2j} + (1-\lambda)\partial V_2/\partial n_{2j} = \lambda\partial g_1/\partial n_{2j} + (1-\lambda)\partial g_2/\partial n_{2j} - c'(n_{1j} + n_{2j}) = 0. \qquad \text{(IV.7)}$$

The contract curve connects the type optimal allocations, as shown in Figure 2.

V. Equilibrium clusters

A. Migration and Nash equilibrium

Following Henderson (1974, 1988) and many others, we assume that equilibrium city sizes are determined by costless individual migration by workers.[11] In this case, the conditions that must be satisfied in order for an allocation of J clusters with populations $(n_{1j},n_{2j},\ldots,n_{Ij})$ to constitute a Nash equilibrium are:

1. Utility maximization. Taking the utility levels offered in other clusters as given, there must be no cluster that offers a higher level of utility to a worker of any type.
2. Stability. There must be no incentive for workers to move.
3. Adding up. Equilibrium allocations must satisfy the aggregate population constraints.

Although these conditions are standard, a few comments about their application in this multi-industry setting are warranted. Utility maximization implies that every equilibrium allocation that contains positive numbers of type-i workers must offer them the same level of utility. In particular, if $(n_{1j},n_{2j},\ldots,n_{Ij}) > 0$ and $(n'_{1j},n'_{2j},\ldots,n'_{Ij}) > 0$ are both equilibria, then it must be the case that $V_i(n_{1j},n_{2j},\ldots,n_{Ij}) = V_i(n'_{1j},n'_{2j},\ldots,n'_{Ij})$ for all $i \in \{1,2,\ldots,I\}$. Utility maximization also implies that all workers must be at least as well off in an equilibrium cluster as they would be in autarky: if $(n_{1j},n_{2j},\ldots,n_{Ij})$ is an equilibrium allocation, then $V_i(n_{1j},n_{2j},\ldots,n_{Ij}) \geq V_i(0)$, for all i

[11] An alternative approach, also pioneered by Henderson (1974, 1988), supposes that the spatial development of the economy is directed by large agents -- governments or land developers -- who provide infrastructure and offer contracts to attract workers to locations to maximize land values or profits. We discuss how this approach could be applied to our model later in the paper.

$\in \{1,2,\ldots,I\}$. Let A_i denote the set of allocations that are at least as good as autarky for a type-i worker, that is $A_i = \{(n_{1j},n_{2j},\ldots,n_{Ij}): V_i(n_{1j},n_{2j},\ldots,n_{Ij}) \geq V_i(0)\}$, $i \in \{1,2,\ldots,I\}$. A Nash equilibrium must lie in the intersection $\cap_i A_i$.

Stability implies that all equilibrium allocations $(n_{1j},n_{2j},\ldots,n_{Ij})$ must satisfy $\partial V_i(n_{1j},n_{2j},\ldots,n_{Ij})/\partial n_{ij} < 0$ for all $i \in \{1,2,\ldots,I\}$. To see this, note that if a type-i worker moves between the equilibrium allocations $(n'_{1j},n'_{2j},\ldots,n'_{Ij}) > 0$ and $(n_{1j},n_{2j},\ldots,n_{Ij}) > 0$, the resulting change in utility is $V_i(n_{1j},n_{2j},n_{i-1j},n_i+dn_{ij},n_{i+1j},\ldots,n_{Ij}) - V_i(n'_{1j},n'_{2j},\ldots,n'_{Ij})$. By the equal utility condition, this can be written $V_i(n_{1j},n_{2j},n_{i-1j},n_i+dn_{ij},n_{i+1j},\ldots,n_{Ij}) - V_i(n_{1j},n_{2j},\ldots,n_{Ij}) \approx \partial V_i(n_{1j},n_{2j},\ldots,n_{Ij})/\partial n_{ij}$. Thus, if $\partial V_i(n_{1j},n_{2j},\ldots,n_{Ij})/\partial n_{ij} < 0$ for all $i \in \{1,2,\ldots,I\}$, and for all equilibrium allocations $(n_{1j},n_{2j},\ldots,n_{Ij})$, then it is impossible for a worker to increase her utility by moving. In other words, $\partial V_i(n_1,n_2,\ldots,n_I)/\partial n_i < 0$ ensures that every allocation that satisfies the equal utility condition is stable with respect to individual migration.[12] Let $S_i = \{(n_{1j},n_{2j},\ldots,n_{Ij}): \partial V_i(n_{1j},n_{2j},\ldots,n_{Ij})/\partial n_i < 0\}$, $i \in \{1,2,\ldots,I\}$ denote the set of allocations that are stable for a type-i worker. A Nash equilibrium must lie in the intersection $\cap_i S_i$.

The following result is fundamental:

Proposition 2: Suppose, as in Assumption 2, that adding another industry i worker has a larger impact on productivity in industry i than in industry k ($\partial g_i/\partial n_{ij} > \partial g_k/\partial n_{ij}$, $k \neq i$), for any employment vector $(n_{1j},n_{2j},\ldots,n_{Ij}) > 0$. Then a Pareto efficient allocation cannot be a Nash equilibrium.

Proof: Let $(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})$ be Pareto efficient. From (IV.5), $(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})$ satisfies $\sum_i \lambda_i \partial V_i(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_k = \sum_i \lambda_i(\partial g_i(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_k) - c'(N^j) = 0$, $\lambda_i > 0$, for all $k \in \{1,2,\ldots,I\}$. Stability requires $\partial V_i(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_i = \partial g_i(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_i - c'(N^J) < 0$ for all $i \in \{1,2,\ldots,I\}$. Assumption 2 states that $\partial g_i(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_i > \partial g_k(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_i$, $k \neq i$, and this in turn implies that if $\partial V_i(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_i < 0$, then $\partial V_k(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_i < 0$, $k \neq i$. Thus, Assumption 2 and stability imply that $\sum_i \lambda_i \partial V$-

[12] Stability of Nash equilibrium does not impose restrictions on $\partial V_i/\partial n_{kj} < 0$, $k \neq i$, since group deviations are not allowed. For discussions of dynamic stability in city systems see Anas (1992), Henderson and Becker (2000) and Tabuchi and Zeng (2004).

$_i(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})/\partial n_k < 0$, $\lambda_i > 0$, for all $k \in \{1,2,\ldots,I\}$, contradicting (IV.5) and the assumption that $(n^{PE}_{1j},n^{PE}_{2j},\ldots,n^{PE}_{Ij})$ is Pareto efficient. QED

It is striking that a Pareto efficient allocation cannot be an equilibrium in this general multi-industry setting. This result is in marked contrast with results from single industry models (see Henderson and Becker (2000)). In those models, with only one industry or worker type, an allocation in which all workers reside in identical cities of optimal size is, under suitable assumptions, stable with respect to individual migration. The instability that is at the core of Proposition 2 arises from differences in the optimal composition of local employment by industry, and is thus unique to our multi-industry approach. When type optimal cities differ in employment composition, a Pareto efficient allocation that is in the interior of the contract surface places all workers in cities that are not type optimal. Assumption 2 implies that, from such a position, all workers have an incentive to move. It is important to note that the conditions of Assumption 2 are stronger than is required for Proposition 2 to hold. All that is required is that a Pareto efficient allocation be unstable for one type of worker. To guarantee this, it would be sufficient to assume that $\partial g_i/\partial n_{ij} > \partial g_k/\partial n_{ij}$, $k \neq i$, for at least one industry i, rather than all for all $i \in \{1,2,\ldots,I\}$.[13]

Figure 3 illustrates Proposition 2 for I = 2. To help visualize the results, it is useful to note that (IV.1) implicitly defines a set of loci $n^{*i}_{kj}(n_{1j},n_{2j},\ldots,n_{k-1j},n_{k+1j},\ldots,n_{Ij})$, i, k, $\in \{1,2,\ldots,I\}$ such that $\partial V_i/n_{kj} = 0$. The type-i optimal allocation occurs at the intersection of these loci for a given i. Assumptions 1 and 2 imply $n^{*1}_1(0) > n^{*1}_2(0)$, $n^{*1}_1(0) > n^{*2}_1(0)$, $n^{*2}_2(0) > n^{*2}_1(0)$ and $n^{*2}_2(0) > n^{*1}_2(0)$. Figure 3 shows all four loci, in their correct orientations, the type optimal allocations and a contract curve for this example. Any point on $n^{*1}_1(n_2)$ satisfies $\partial V_1/\partial n_1 = 0$. To the left of $n^{*1}_1(n_2)$, it must be the case that $\partial V_1/\partial n_1 > 0$ by concavity. Similarly, concavity implies that all points beneath the locus $n^{*2}_2(n_1)$ satisfy $\partial V_2/\partial n_2 > 0$. In Figure 3 the stable region $\cap_i S_i$ corresponds to the shaded area to the right of the intersection of $n^{*1}_1(n_2)$ and $n^{*2}_2(n_1)$. Proposition 1 implies that the contract curve connecting the type optimal allocations lies in the region where $\partial V_1/\partial n_1 > 0$ and $\partial V_2/\partial n_2 > 0$, as shown in the figure. All points on the contract

[13] The assumption that every industry produces the numeriare simplifies the proof of Proposition 2 since output prices do not influence the relationship between productivity, wages and utility. If output prices differ by industry, then Proposition 2 would require, at a minimum, $p_i\partial g_i/\partial n_{ij} > p_k\partial g_k/\partial n_{ij}$, $k \neq i$, for at least one industry i.

curve are unstable with respect to individual migration, and, consequently, a Pareto efficient allocation cannot be a Nash equilibrium.

B. Symmetric equilibria

The previous section focused on the implications of utility maximization and stability for the characteristics of Nash equilibrium cluster sizes and compositions. This section considers the consequences of the adding up condition. We begin with symmetric equilibria in which all clusters have the same size and composition; asymmetric equilibria are considered below.

Recall that the adding up constraint, in its most general form, requires $\sum_j n_{ij} = N_i$, where N_i is the given aggregate number of type-i workers in the economy. In a symmetric allocation with J identical clusters, the adding up constraint implies $Jn_i = N_i$ for all $i \in \{1,2,\ldots,I\}$. Thus, symmetric allocations that satisfy adding up must lie on the *ratio vector*:

$$(n_1, n_2,\ldots,n_I) = (1/J)(N_1,N_2,\ldots,N_I). \qquad \text{(V.1)}$$

The intersection of the ratio vector from (V.1) with the contract surface from (IV.5) defines a symmetric constrained Pareto efficient allocation $(n^{CPE}_1, n^{CPE}_2,\ldots,n^{CPE}_I)$ for this economy. For I = 2, the number of type-1 and type-2 workers in a symmetric allocation must lie on the *ratio line*

$$n_2 = (N_2/N_1)n_1. \qquad \text{(V.2)}$$

In this case, (n^{CPE}_1, n^{CPE}_2) lies at an intersection of the ratio line from (V.2) and the contract curve from (IV.6) and (IV.7), as shown in Figure 3.[14] Proposition 2 implies that a constrained Pareto efficient allocation cannot be a Nash equilibrium.

A symmetric Nash equilibrium must lie in the intersection of the at least as good as autarky sets $\cap_i A_i$, the stable region $\cap_i S_i$, and the ratio vector in (V.1). For I = 2, the set of symmetric Nash equilibria corresponds to the portion of the ratio line that is at least as good as autarky for both types of workers and lies to the right of the intersection of $n^{*1}_1(n_2)$ and $n^{*2}_2(n_1)$. This is shown as the bold part of the ratio line in the Figure 4.

[14] Such an allocation will exist provided $n^{*2}_2/n^{*2}_1 > N_2/N_1 > n^{*1}_2/n^{*1}_1$ – that is, provided the ratio line intersects the contract curve between the type optimal allocations.

Figure 4 illustrates several important features of symmetric Nash equilibria in this multi-industry setting. First, a symmetric Nash equilibrium will exist if, for some J, the intersection of the least as good as autarky sets $\cap_i A_i$, the stable region $\cap_i S_i$, and the ratio vector is non-empty. Existence cannot be guaranteed in the general case, but it can established in particular cases, including the example presented in the Appendix. Second, when an equilibrium exists, it will generally not be unique. Ignoring integer issues, all points along the bold portion of the ratio line in Figure 4 are symmetric Nash equilibria for some J. Third, and perhaps most important, Figure 4 clearly illustrates the nature of the inefficiency that arises from individual migration in this model. Moving up the ratio line in Figure 4 beyond its intersection with the contract curve is analogous to increasing the population of a single industry city beyond its optimal size. In the symmetric case, the mix of types in any cluster is fixed along the ratio line by the aggregate population constraints, and hence is the same in efficient and equilibrium allocations. However, symmetric Nash equilibrium clusters contain excessive numbers of workers of both types – as in single-industry models, individual migration leads to cities of excessive scale.

VI. Mixed and specialized cities

Section V considered coagglomeration in a system of symmetric cities. We now consider the possibility of asymmetric equilibria, beginning with the most natural sort of asymmetric system, one involving completely specialized cities. Section VII goes on to consider a more expansive universe of asymmetric cities where cities are defined by the matching of various industries. In these sections, the focus of the welfare analysis will move from the *scale* of agglomeration to its *composition*.

A. Optimal specialized cities

In considering specialized cities, it is helpful to begin with an optimum program. Let $V_i^k(n_{kj}) = V_i(0,0,\ldots n_{kj},\ldots, 0,0)$ denote the level of utility that a type-i worker obtains in completely specialized city-j that includes only workers of type-k (i.e., one where $n_{k'j} = 0$ for $k' \neq k$). In this notation, $V_i^i(n_{ij})$ denotes the utility that a type-i worker obtains in a city containing only its own type. While the general utility function $V_i(n_{1j},n_{2j},\ldots,n_{Ij})$ discussed in the last section was new in its consideration of coagglomeration, $V_i^i(n_{ij})$ is familiar. It gives the utility possibilities frontier for a particular activity (type of worker) as a function of the scale of that activity. This sort of utility possibilities curve is central to the Henderson (1974) model of a system of cities. The assumptions made previously mean that $V_i^i(n_{ij})$ has an inverted U-shape, as in Henderson and successors.

Let the *type-i optimum specialized city* population be denoted by n_i^S = argmax $V_i^i(n_{ij})$. The relationship of this to the previously defined concept of a type-optimum city is most easily demonstrated by referring Figure 3. Recall that the locus $n_i^{*i}(n_k)$, $k \neq i$, denotes the population of type-i workers that maximizes a type-i worker's utility taking n_k as given. The type-i optimal specialized city in this case is $n_i^s = n_i^{*i}(0)$, the intercept on the n_i axis of the locus of $n_i^{*i}(n_k)$. Thus, by construction, there always exist allocations involving coagglomeration that a type-i worker would prefer to even the best possible specialized city. In particular, the unconstrained type-i optimal city is preferred by a type-i worker to the specialized city represented by n_i^S.

B. Equilibrium specialized cities

An equilibrium system of completely specialized cities must satisfy the conditions discussed in Section V: utility maximization, stability, and adding up. Adding up determines the number of specialized cities. Denote the set of cities containing a type-i worker by J^i. Then, adding up implies that the cardinality of J^i satisfies, $|J^i| = N_i/n_i^0$, where n_i^0 is common population level implied by equal utility (see below).[15] Stability in this case requires $\partial V_i^i(n_{ij})/\partial n_{ij} < 0$ for all $j \in J^i$, or, equivalently, $(0,0,\ldots,n_{ij},\ldots, 0,0) \in S_i$. Graphically, it means that the city population must be on the downward sloping part of $V_i^i(n_{ij})$.

Utility maximization has three implications in this setting. First, as above, it implies that all cities containing positive numbers of type-i workers must offer them the same level of utility. This in turn implies that all specialized cities containing type-i workers must be the same size (hence n_i^0 in adding up condition). The second implication of utility maximization is that all workers must be at least as well off in an equilibrium city as they would be in autarky: thus, $V_i^i(n_{ij}) \geq V_i(0)$, or, equivalently, $(0,0,\ldots n_{ij},\ldots, 0,0) \in A_i$. The third and new implication of utility maximization is that the level of utility of a type-i worker in a specialized city must be at least as great as it would be in any populated city that does not contain type-i workers. This means that in addition to being preferred to autarky, a worker must prefer a specialized city made up of its own type to being the first worker in another type of specialized city. Formally, this requires that

[15] As above, we are assuming away integer issues here.

in order for the vector of specialized city populations $(n_1,n_2,...,n_I)$ to be an equilibrium, we must have

$$V_i^i(n_i) \geq V_i^k(n_k) \text{ for all } k \neq i \text{ .} \qquad \text{(VI.1)}$$

If industries have strong complementarities, (VI.1) may restrict the range of equilibrium city sizes, and in particular, may preclude populations that will give only the autarky utility level. This point is illustrated in Figure 5. The figure shows $V_i^i(n_i)$ and $V_i^k(n_k)$ for some i and k. Let n_i^A be the specialized city population that is equivalent to autarky: $V_i^i(n_i^A) = V_i^i(0)$. Then, if there were only one industry or worker type, any population between n_i^S and n_i^A that satisfies the adding up constraint could be an equilibrium, since all such points are stable with respect to individual migration. However, if there are many industries or worker types, and a specialized city made up of type-k workers has population n_k^S, then populations between n_i^C and n_i^A (where $V_i^i(n_i^C) = V_i^k(n_k^S)$) in Figure 5 cannot be part of an equilibrium: in this range, a type-i worker would rather join the type-k city.

In order to assess the welfare properties of specialized equilibria, we will make a strong assumption that ensures that (VI.1) holds for all $n_i \in [n_i^S,n_i^A]$:

Assumption 3: $V_i(n_i^A) \geq V_i^k(n_k^S)$ for all $k \neq i$. (VI.2)

Assumption 3 means that complementarities are weak enough that all types of workers prefer a specialized city with their own type with population large enough to reduce utility to the autarky level to the type optimal specialized city for any other type.

This leads directly to the following result:

Proposition 3: When there are weak complementarities as defined in Assumption 3, there exist a range of specialized city Nash equilibria where for any worker type i, $n_i \in [n_i^S,n_i^A]$.

Proof: Suppose that populations are at the type-i optimum specialized city levels, n_i^S, for all i. This satisfies equal utility since all city sizes are the same, and the number of cities can be adjusted to satisfy adding up. Stability holds as well. By Assumption 3, no worker prefers another worker's specialized city, so this allocation meets all the equilibrium conditions. Suppose that the allocation involved cities with populations $n_i^0 \in (n_i^S, n_i^A]$. As long as the city sizes are all equal, equal utility, adding up, and stability all hold. Since $V_i(n_i^0) < V_i(n_i^A)$, no worker again prefers another specialized city, so this allocation is also an equilibrium. QED

Assumption 3 is thus a sufficient condition guaranteeing the existence of a range of completely specialized cities where cities have populations between n_i^S and n_i^A. That there are many sizes of specialized cities consistent with equilibrium is the fundamental migration pathology in its most basic form.

However, we are more concerned with a different sort of welfare effect, specifically whether individual migration always generates coagglomeration when it is efficient. The answer is, that it does not. To see this, consider the case where a specialized equilibrium generates autarky utility levels. Such an equilibrium is Pareto inferior to an equilibrium with mixed cities where any type achieves utility greater than under autarky. As long as the set of equilibria given by the intersection of the better-than-autarky sets $\cap_i A_i$, the stable region $\cap_i S_i$, and the ratio vector is non-empty, there exists a mixed city equilibrium. As long as the intersection is non-degenerate (i.e., as long as the segment of the ratio line intersection the stable sets includes multiple allocations that are better than autarky for all types), there exists a mixed equilibrium that is superior to autarky. Since there exists a specialized equilibrium that gives all types their autarky payoffs, this means that there exists a specialized equilibrium that is inferior to a mixed equilibrium where no worker would move from the specialized city of its own type to a specialized city of another type.

The source of this inefficiency is the weakness of individual migration in generating efficient cluster composition. Assumption 3 means that no worker is willing to be the first to give up the benefits of locating with other workers of the same type in order to locate with some other type of worker. This is sufficient to sustain a specialized-city equilibrium, even when a coagglomerated equilibrium is globally superior. All that is required is that the first worker not be willing to move.

This result sharply challenges conventional wisdom. It is very common to explain the composition of cities and business clusters by a sort of “just so” story. When agglomeration economies fail to extend beyond industry boundaries (localization economies) cities will tend to specialize. When agglomeration economies do extend beyond industry boundaries (urbanization economies) cities will be diverse. Thus, Chinitz (1961) attributes the specialization of Pittsburgh to localization economies that are confined to the steel industry, and the diverse economy of New York to more general urbanization economies. A cogent statement of this view is found in Henderson (1988, p. 78):

> Specialization occurs if there are no production benefits or positive externalities from locating two different industries in the same place … Industries that use each other's inputs, a common labor force, or a common public good or intermediate input … will tend to locate together.

This section's analysis shows that the just so approach to clustering is misleading. In particular, the analysis shows that coagglomeration of related industries can fail to arise even when such clustering would be welfare enhancing.

VII. Migration and cluster composition

The previous section showed that specialized cities may persist even when diversity would be superior. This section completes the dismal story by showing that industries may cluster even when they do not benefit from each other. To do this we will consider the general problem of the matching of industries in cities when workers are perfectly mobile. As noted above, the informal welfare analysis in the literature tends to treat this matching as being optimal, where industries coagglomerate to their benefit. This section will show that when cluster formation is guided by individual migration, this need not be so.

A. Inefficient partitions

We will illustrate the possible inefficiency of cluster composition by analyzing a particular form that the matching of worker types to clusters can take. Specifically, we will consider allocations that are based on a partition of the set of worker types into $K < I$ disjoint and mutually exhaustive subsets P_k such that $\cup_k P_k = \{1,2,\ldots,I\}$. Assume that $n_i > 0$ if and only if i is an element of P_k. In this setup, industries i and j are coagglomerated if and only if i and j are both elements of P_k. For example, in the partition $P_1 = \{1,2,\ldots,i\text{-}1\}$ and $P_2 = \{i,i+1,\ldots,I\}$ industries 1,2,…,i-1 are coagglomerated, as are industries i,i+1,…,I. We use the partition to define a finite number of city types T, where a city type is a vector of populations for the industries that are coagglomerated in the partition. For example, all type 1 cities from the example given above would have composition $(n_{1j},n_{2j},\ldots,n_{i-1j},0,0,\ldots 0)$. As before, we confine attention to symmetric cities whose numbers and populations satisfy the adding up constraint. For example, letting m_1 denote the number of symmetric type 1 cities in the example above, adding up implies that $n_{ij} = N_i/m_1$ for i in P_1 and 0 otherwise. We will refer to the resulting allocation of workers to cities as a *symmetric partition*.

We will now make a strong assumption about technology. Specifically, we will suppose that for each worker type, some agglomeration with its own type is essential. Formally, the assumption is:

Assumption 4: $V_i(n_{1j},n_{2j},\ldots, n_{i-1j},0,n_{i+1j},\ldots,n_{Ij}) < V_i(0)$, for any $(n_{1j},n_{2j},\ldots, n_{i-1j},0,n_{i+1j},\ldots,n_{Ij}) \geq 0$.

Assumption 4 means that a worker locating with no own-type workers has a lower utility level than in autarky.

This gives the following:

Proposition 4: Suppose, as in Assumption 4, that own-type workers are essential. Then any symmetric partition that satisfies stability and weakly dominates autarky for all worker types is a Nash equilibrium.

Proof: An equilibrium must satisfy the conditions discussed above: utility maximization, stability, and adding up. Utility for a given worker type is equal in every inhabited city by symmetry. Stability is satisfied by hypothesis. Adding up is met by the definition of a partition. The remaining requirement is that no worker of any type be able to benefit by moving to another city. By the definition of partition, the alternate city would have no own type workers. By Assumption 4, this would give a utility level lower than autarky, which is lower than that of the symmetric partition, by hypothesis. QED

Proposition 4 shows that when the presence of own type workers is essential, the only restrictions on equilibrium matching are stability, equal utility, and better than autarky. This places very weak restrictions on which types of workers (industries) will coagglomerate with which other types. Thus, a worker is willing to tolerate sharing a city with other workers from whom no benefit is derived. For example, the San Francisco Bay Area's oil refining workers coagglomerate with it software engineers, even though there are probably few direct complementarities between them. This could be explained by the strength of own-industry agglomeration effects, as in Assumption 4.

The forces at work become more transparent when one considers an even simpler example. Suppose that there are four worker types (I = 4) and two occupied cities (J = 2). Suppose that all the type 1 and 2 workers coagglomerate in city 1, while the type 3 and 4 workers coagglomerate in city 2. This is a partition with two types of cities. Suppose that stability holds and that the partition gives exactly the autarky level of utility to all four types:

$$w_1(N_1,N_2,0,0) - c(N_1+N_2) = V_1(0), \quad \text{(VII.1a)}$$
$$w_2(N_1,N_2,0,0) - c(N_1+N_2) = V_2(0), \quad \text{(VII.1b)}$$
$$w_3(0,0,N_3,N_4) - c(N_3+N_4) = V_3(0), \quad \text{(VII.1c)}$$
$$w_4(0,0,N_3,N_4) - c(N_3+N_4) = V_4(0). \quad \text{(VII.1d)}$$

The allocation defined in (VII.1a) - (VII.1d) is an equilibrium.

Consider an alternate allocation of worker types to cities with worker types 1 and 3 in one city and worker types 2 and 4 in the other. Suppose that this allocation gives a strictly greater level of utility to all four types:

$$w_1(N_1,0,N_3,0) - c(N_1+N_3) > V_1(0), \quad \text{(VII.2a)}$$
$$w_2(0,N_2,0,N_4) - c(N_2+N_4) > V_2(0), \quad \text{(VII.2b)}$$
$$w_3(N_1,0,N_3,0) - c(N_1+N_3) > V_3(0), \quad \text{(VII.2c)}$$
$$w_4(0,N_2,0,N_4) - c(N_2+N_4) > V_4(0). \quad \text{(VII.2d)}$$

If one were to begin with the inferior partition of 1 with 2 and 3 with 4, no individual would be willing to move to a city with the better partner industry by Assumption 4. This is true even beginning from a very bad match, with all types enjoying only autarky utility. As above, a worker may be willing to tolerate coagglomeration with other workers from whom no benefit is derived, if own-industry effects are sufficiently strong.

The fundamental issue is that individual migration is a weak instrument of competitive discipline. The analysis in this section shows that when the attraction to the own-type is strong enough, inferior matches can persist. Assumption 4 is thus a sufficient condition for bad matches to persist in equilibrium. It is easy to see that weaker conditions can also preclude efficient matching. For instance, it is not necessary that a cluster with none of a worker's own type give the worker lower utility than autarky. It is only necessary that it give lower utility than in the cluster containing own type workers from which the worker is departing.

This inefficient matching is parallel to Bewley's (1981) analysis of sorting in a system of local governments. He presents a number of situations where efficiency fails for local public goods, despite the presence of mobility that might be thought to allow the sort of shopping that takes place in Tiebout (1956). Bewley's general result is that efficiency requires more than simply allowing households to "vote with their feet." It requires also, among other things, entrepreneurial incentives on the part of public goods providers.[16]

B. Dynamics and institutions

This section shows inefficient coagglomeration between industries to be consistent with equilibrium. This unfortunate possibility is demonstrated by characterizing inefficient equilibrium allocations that cannot be upset by individual migration decisions. It is natural to consider how such an equilibrium might arise. The standard dynamic story that is told in the traditional single industry systems of cities literature (Henderson (1974)) involves one-by-one location decisions. In this heuristic dynamic, workers arrive one at a time and migrate to the city that offers the highest utility. The relationship of utility to population is assumed to have an inverted U-shape. Workers then fill up the first city to the point that the city offers maximum utility. They continue to migrate to the first city beyond this point, since it offers greater utility than autarky (i.e., in a city of size zero). This process continues until utility in the city is driven

[16] For extensions of Bewley's normative analysis, see Ellickson et al (1999) and the many papers surveyed by Scotchmer (2002). As noted by Scotchmer, the literature contains both welfare theorems, showing that a price system can lead to efficiency, as well as parallel results showing that constraints on pricing lead to failures of efficiency. Our analysis of co-agglomeration obviously falls in the latter category.

down to the autarky level. This dynamic story supports the most inefficient of the many possible inefficient equilibria in a one-type system-of-cities model.

The parallel here would be to assume that workers arrive one-at-a-time. If their arrival occurred in proportion to population shares, then the outcome would be very similar to the one-type outcome. Workers would fill the first city to the point where the utility in the mixed city equals the level under autarky. This would give a mixed city even if such a city were Pareto inferior to some other pattern of coagglomeration or to specialization.

This is not the only possible dynamic. One could instead consider a regime where workers arrive according to a multinomial random process, with probabilities equal to population shares. As long as the migration continued to be one-by-one, the potential for inefficiency would remain. Another possibility is to assume that the workers arrive instead in lumps, where the lumps correspond to firms. In this case, if the firms were large or in the multinomial model if there were a run of firms of a particular type, it could break the pattern of mixed cities that has arisen thus far.

Although this one-by-one dynamic of the urbanization of a national economy has a lot of appeal, it is not the only dynamic process at work. Another involves changes in the nature of agglomeration economies. These changes certainly have occurred in the past, and they seem to be occurring with great frequency now. For instance, it was once the case that a firm's back office activities (e.g., accounting and billing) needed to be located near the firm's front office. Improvements in communications technology have changed this relationship, and one now observes back offices located in different cities than their affiliated front offices. One way to interpret the demonstrations of sustainable inefficiencies is as capturing circumstances that could have arisen as part of the process of technological change and that would not have upset the existing pattern of coagglomeration. So the inefficiency may not be arising as a result of active migration but instead as a result of a failure of adaptation. While we are not able to provide direct evidence of this failure, it does seem to be quite plausible.

As noted previously, the failure to realize efficient coagglomeration in this analysis depends on the assumption that individual migration is the only force that might correct an inefficient pattern of coagglomeration. An alternative approach (Henderson, (1974)) is to suppose that cities are formed by "developers." In some conceptions, developers are private agents seeking to maximize land value. In others, they are city or regional governments attempting to maximize local welfare. These agents must have the power to exclude workers from the cities that they manage and the ability to offer incentive compatible city contracts (see Helsley and Strange (1990) for details), and this can have profound impacts on the realization of efficient allocations of population. In Section III's model, the ability to restrict migration could allow realization of the constrained Pareto efficient allocation. In this section's analysis, developers could engineer the group deviations and exclusions that would realize efficiency.

So the question becomes: are there real world institutions that correspond to such powerful city developers? Our answer is: no. Ascribing entrepreneurial incentives to city governments seems to be at odds with actual city government behavior.[17] Even if local governments were fully entrepreneurial, the fragmentation of local government means that exclusion of the sort required for efficiency is not possible. Private developers are rarely large enough to be reasonably seen as controlling the development of entire cities.[18]

VIII. Empirical implications

One of the most important questions in the field of urban economics is: what are the sources of agglomeration economies? There are many ways that this question has been answered. One approach has been to examine the forces that lead to the agglomeration of particular industries or groups of industries. Another is to look in isolation at individual agglomerative forces. These approaches have demonstrated persuasively the existence of a number of individual agglomeration economies in a range of industries. A third approach is to look at many agglomeration economies and many industries together, and to look for evidence of the relative importance of various microfoundations. This approach is taken by Audretsch and Feldman (1996) and Rosenthal and Strange (2001), who regress measures of the spatial concentration of industries on proxies for various Marshallian localization economies. These approaches are based the idea that observed location patterns reflect the benefits that can be realized from agglomeration.

---

[17] See also the survey by Scotchmer (2002). She writes (p. 2002) that "…the idea of a price-taking equilibrium is rather far afield from the way local public economies operate, and this is why I consider club theory to be a motivator for the subject of local public economics, but not the subject itself."

[18] See Helsley and Strange (1997) for a formal model of limited developers and institutional details, including the few situations where developers have come closest to total control. Furthermore, as shown in Helsley and Strange (1994), if developers were powerful enough to compete as firms, the nature of urban development means that the competition would be imperfect, leading to an equilibrium that falls short of first best.

A particularly creative way to look for the relative importance of various microfoundations is carried out in Ellison et al (2010). Their approach is to infer the importance from patterns of coagglomeration. Suppose that industries that tend to coagglomerate also tend to draw from the same patents. This suggests that knowledge spillovers are important. Suppose instead that industries that tend to coagglomerate tend to employ the same sorts of manufactured inputs. This suggests that input sharing is important, and so on. Ellison et al (2010) implement this idea by regressing coagglomeration between industry pairs on indices capturing the industries' tendencies to draw from related resources. This approach is based on the idea that observed location patterns reflect the benefits that can be realized from coagglomeration.

The previous two sections show two important sources of inefficiency in the composition of clusters. As implied by Proposition 3, there may be unrealized coagglomeration benefits. In this case, a worker from a particular industry is not attracted to another industry *per se* but instead is attracted to workers from their own industry. In this case, one may fail to observe in equilibrium the most efficient possible coagglomeration. For example, knowledge spillovers may be potentially very valuable, but may not be observed in equilibrium. As implied by Proposition 4, it is possible that there be inefficient diversity. A worker from a particular industry may choose a big and diverse city not for its diversity *per se* but instead because a big city has a lot of activity in the same industry. The agglomeration benefit is by virtue of size and not by virtue of composition. In this unfortunate situation, one could see workers choosing to coagglomerate even when they do not share the same orientations with respect to ideas, people, or goods.

IX. Conclusion

Traditional models of systems of cities focus almost entirely on the polar cases completely mixed and fully specialized cities. There has been little consideration of coagglomeration. In the absence of formal analysis, the folk wisdom about coagglomeration is that it is efficient. If there are localization economies only, then cities will be specialized. Only if there are sufficiently strong urbanization economies will diverse cities emerge.

This paper challenges the idea that coagglomeration can be expected to be efficient. In some situations, industries will coagglomerate inefficiently. In others, there will be unrealized coagglomeration that could have enhanced efficiency. These results are important as extensions to the theory of systems of cities and as bases for urban policy analysis. They are also important for shedding light on what can be learned about the microfoundations of agglomeration economies from observed patterns of coagglomeration.

There is one additional point about coagglomeration that this paper has not considered, and it bears making here. In much of the analysis, we have employed a reduced form relationship between agglomeration and productivity. The relationship is consistent with various microfoundations of agglomeration economies. The inefficiency results arise from the weakness of migration as a disciplining mechanism regardless of the agglomeration forces at work.

In fact, the likelihood of inefficiency is even greater. While some of the benefits of agglomeration arise as a result of the essentially accidental contact of nearby agents, most agglomeration benefits require some sort of action in order for benefits to accrue. In the case of knowledge spillovers, there can be no exchange unless, at a minimum, agents have sufficient levels of human capital to have knowledge to transfer and to learn from other agents. In addition, agents must, in many situations, be willing participants in the exchange (Helsley and Strange (2004)). In the case of input sharing, firms must decide to outsource in order for there to be an increasing return (Helsley and Strange (2007)). For labor market pooling, both employers and workers make decisions that impact the efficiency of agglomeration (Combes and Duranton (2006)). That the benefits of agglomeration depend on what are effectively public goods provision decisions further increases the likelihood of inefficient coagglomeration. If there are no local inputs available because firms have chosen to in-source, then this will further decrease the likelihood of efficient coagglomeration motivated by input sharing.

Clearly, if there are systematic inefficiencies in coagglomeration, then interventions may be beneficial. But who should intervene? Governments have police powers that may allow them to obtain improved matches. However, it is questionable that they have the information needed to recognize the matches and the incentives to correct them if they are recognized. Land developers are likely to have the right incentives, but they are not large enough to internalize all the relevant externalities. The same is true for firms. Given these information issues, this paper's analysis should not be seen as calling for planning solutions to inefficiencies in the system of cities.

References

Abdel-Rahman, Hesham (1994), "Economies of scope and intermediate goods in a system of cities," *Regional Science and Urban Economics* 24, 497-524.

Abdel-Rahman, Hesham (1996), "When do cities specialize in production?" *Regional Science and Urban Economics* 26, 1-22.

Abdel-Rahman, Hesham and Alex Anas (2004), "Theories of Systems of Cities," in Henderson, J.V. and Thisse, J.-F. (Eds.), *Handbook of Urban and Regional Economics, Volume 4*, Elsevier, Amsterdam, 2293-2340.

Abdel-Rahman, Hesham and Masahisa Fujita (1993), "Specialization and diversification in a system of cities," *Journal of Urban Economics* 33, 189-222.

Anas, Alex (1992), "On the birth and growth of cities: Laissez-faire and planning compared," *Regional Science and Urban Economics* 22, 243-258.

Anas, Alex, and Kai Ziong (2003), "Intercity trade and the industrial diversification of cities," *Journal of Urban Economics* 54, 258-276.

Anas, Alex, and Kai Ziong (2005), "The formation and growth of specialized cities: efficiency without developers or Malthusian traps," *Regional Science and Urban Economics* 35, 445-470.

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

Bacolod, Marigee, Bernardo Blum, and William C. Strange (2008), "Skills and the city," *Journal of Urban Economics* 65, 127-135.

Bacolod, Marigee, Bernardo Blum, and William C. Strange (2009), "Urban interactions: soft skills vs. specialization," *Journal of Economic Geography* 9, 227-262.

Bewley, Truman. (1981), "A critique of Tiebout's theory of local public expenditures," *Econometrica* 49, 713-740.

Chinitz, Benjamin (1961), "Contrasts in agglomeration: New York and Pittsburgh," *American Economic Review* 51, 279-289.

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

Combes, Pierre-Philippe and Gilles Duranton (2006), "Labour pooling, labour poaching, and spatial clustering," *Regional Science and Urban Economics* 36, 1-28.

Delgado, Mercedes, Michael. E. Porter, and Scott Stern (2010), "Clusters and Entrepreneurship," *Journal of Economic Geography* 10(4), 495-510.

Duranton, Gilles (2009), "California dreamin': The feeble case for cluster policies," *Review of Economic Analysis*, forthcoming.

Duranton, Gilles and Diego Puga (2001), "Nursery cities," *American Economic Review* 91(5), 1454-1477.

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

Ellickson, Bryan, Birgit Grodal, Suzanne Scotchmer, and William R. Zame (1999), "Clubs and the market," *Econometrica* 67, 1185-1218.

Ellison, Glen and Edward Glaeser (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy* 105, 889-927.

Ellison, Glen, Edward Glaeser, and William Kerr (2010), "What causes industry agglomeration? Evidence from coagglomeration patterns," *American Economic Review* 100, 1195-1213.

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

Glaeser, Edward L., Hedi Kallal, Jose Scheinkman and Andrei Shleifer (1992), "Growth in Cities," *Journal of Political Economy* 100(6) 1126-1152.

Glaeser, Edward L. and Joshua D. Gottlieb (2010), "The wealth of cities: Agglomeration economies and spatial equilibrium in the United States," *Journal of Economic Literature* 47(4), 983-1028.

Glaeser, Edward L. and William R. Kerr (2010), "Clusters of Entrepreneurship," *Journal of Urban Economics* 67(1), 150-168.

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 (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 (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.

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 (1988), *Urban Development: Theory, Fact and Illusion.* New York: Oxford University Press.

Henderson, J. Vernon, Ari Kuncoro and Matt Turner (1995), "Industrial Development in Cities," *Journal of Political Economy* 103(5), 1067-1090.

Henderson, J. Vernon and Randy Becker (2000), "Political economy of city sizes and formation," *Journal of Urban Economics* 48, 453-484.

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

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

Kerr, William R. (2010), "Breakthrough Inventions and Migrating Clusters of Invention," *Journal of Urban Economics* 67(1), 46-60.

Klepper, Steven (2010), "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit," *Journal of Urban Economics* 67(1) 15-32.

Krugman, Paul (1991a), *Geography and Trade*. Cambridge: MIT Press.

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

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

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

Puga, Diego (2010), “The Magnitude and causes of agglomeration economies,” *Journal of Regional Science* 50(1), 203-219.

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 Henderson, J.V. and Thisse, J.-F. (Eds.), *Handbook of Urban and Regional Economics, Volume 4*, Elsevier, Amsterdam, 2119-2172.

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

Saxenian, AnnaLee. (1994), *Regional Advantage: Cultural Competition in Silicon Valley and Route 128* (Harvard University Press: Boston).

Scotchmer, Suzanne. (2002), “Local public goods and clubs,” in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, *Volume 4*, Elsevier, Amsterdam, 1998-2042.

Sundarim, Rangarajan, K. (1996) *A First Course in Optimization Theory.* Cambridge: Cambridge University Press.

Tabuchi, Takatoshi and Dao-Zhi Zeng (2004) “Stability of spatial equilibrium,” *Journal of Regional Science* 44(4) 641-660.

Tiebout, Charles. M. (1956), "A Pure Theory of Public Expenditure," *Journal of Political Economy* 64, 415-424.

Topkis, Donald M. (1998) *Supermodularity and Complementarity.* Princeton, New Jersey: Princeton University Press.

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

Figures

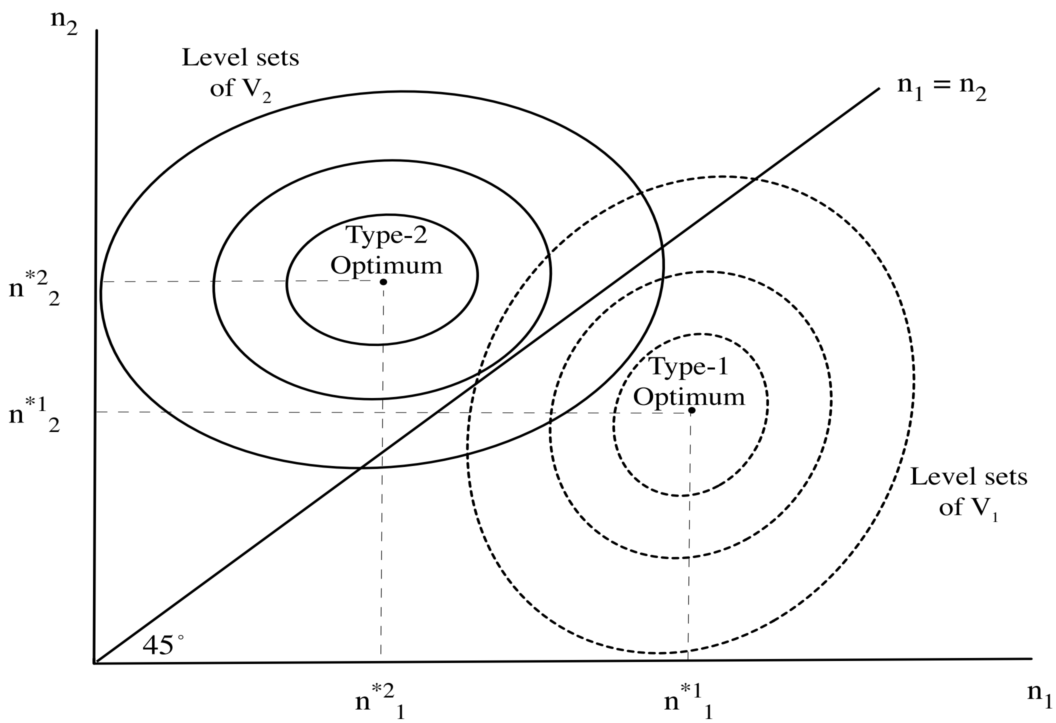

Figure 1: Type optimal clusters and their level sets

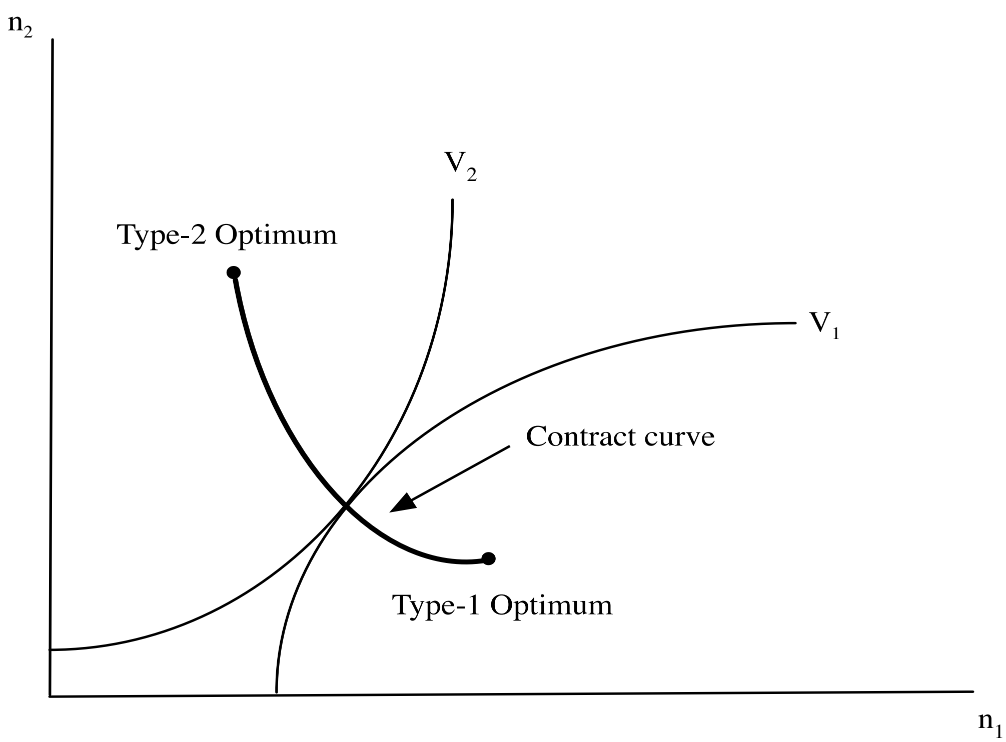

Figure 2: The contract curve

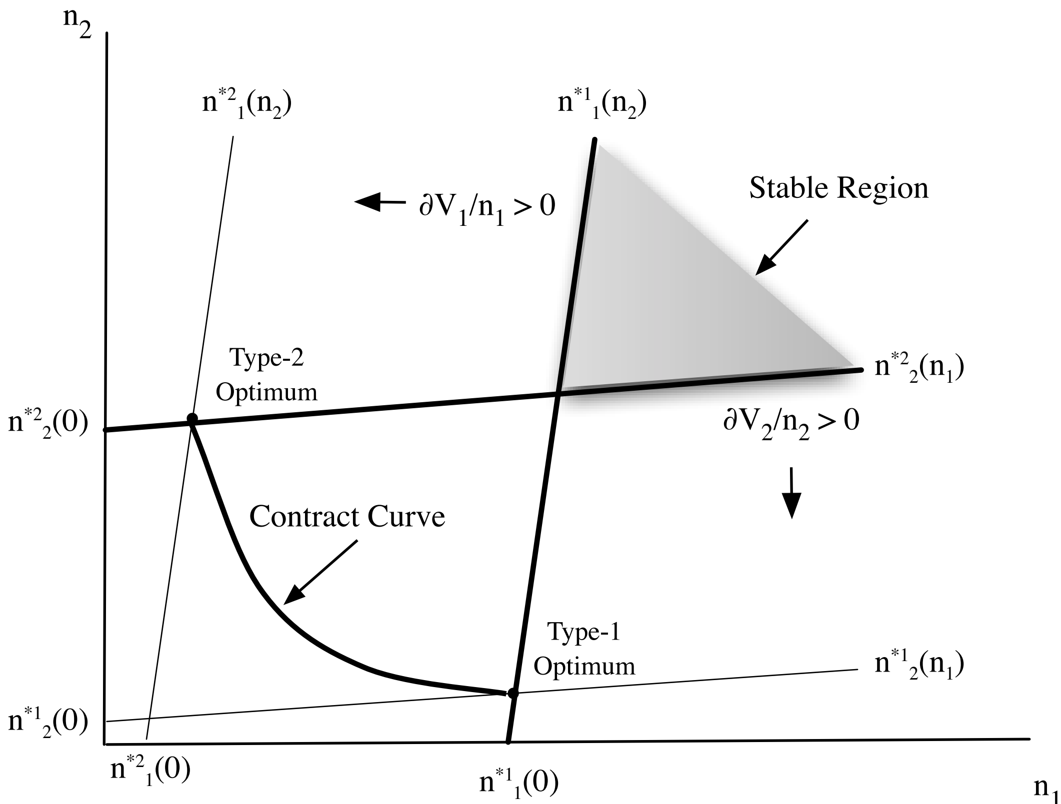

Figure 3: Efficient allocations are unstable

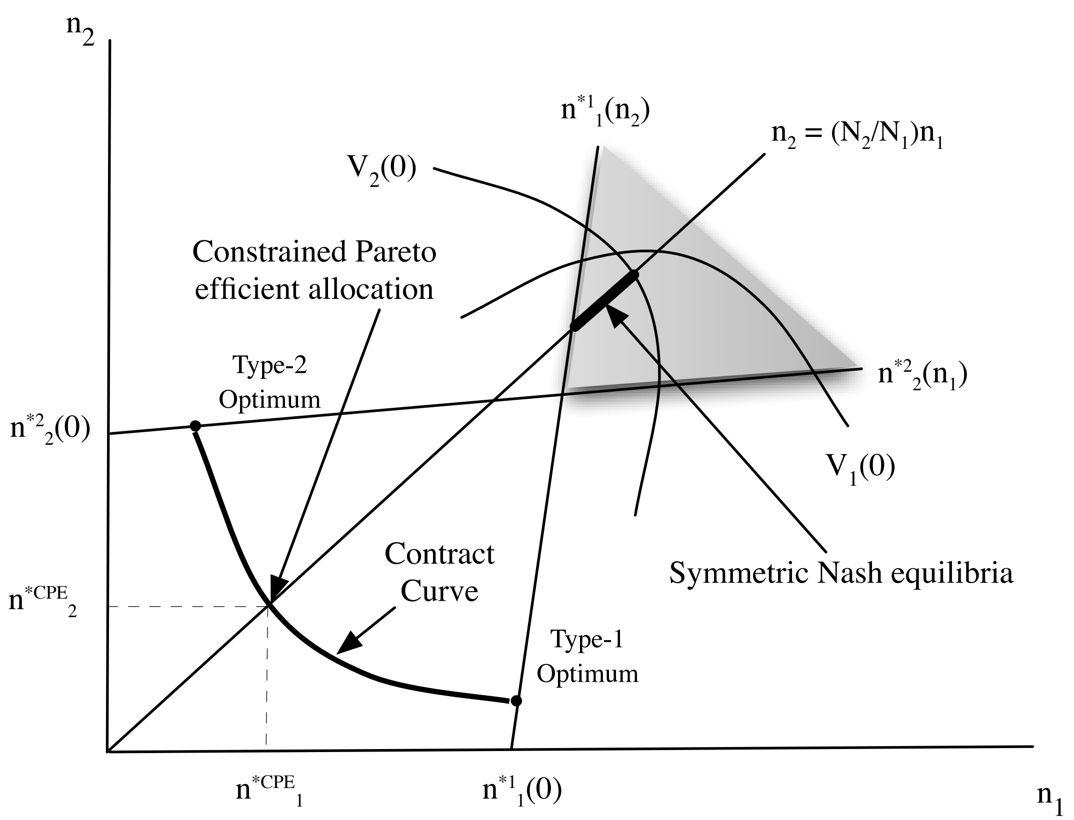

Figure 4: Symmetric Nash equilibria

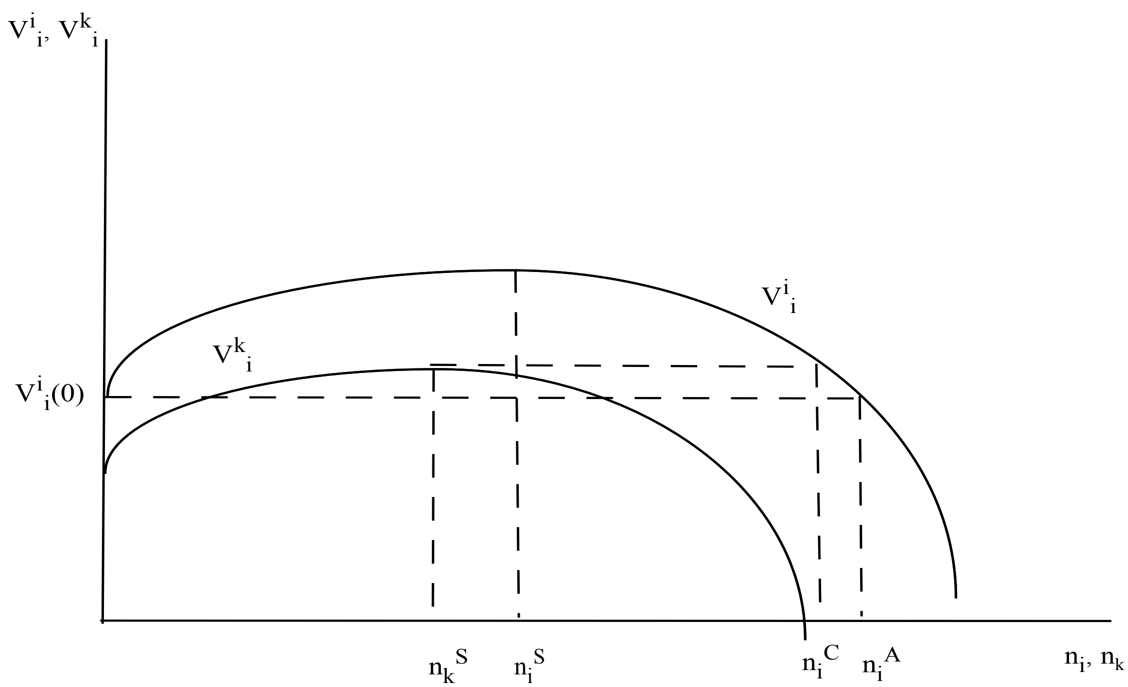

Figure 5: Utility maximization with specialized cities

Appendix - Not for Publication

A specific form $g_i(\cdot)$ that satisfies the assumptions from Section III of the paper is:

$$g_i(n_{1j},n_{2j},\ldots,n_{Ij}) = \sum_s[\theta^i{}_s\alpha n_{sj} - (\beta/2)n_{sj}{}^2] + \gamma\prod_s n_{sj} \qquad \text{(A.1)}$$

where $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\theta^i{}_i > \theta^i{}_s > 1$, and $\theta^i{}_i > \theta^s{}_i > 1$ for all $s \in \{1,2,\ldots,I\}$, $s \neq i$.

In this case,

$$\partial g_i/\partial n_{ij} = \theta^i{}_i\alpha - \beta n_{ij} + \gamma\prod_{s\neq i} n_{sj}, \qquad \text{(A.2)}$$

$$\partial g_i/\partial n_{kj} = \theta^i{}_k\alpha - \beta n_{kj} + \gamma\prod_{s\neq k} n_{sj},\ k \neq i, \qquad \text{(A.3)}$$

$$\partial^2 g_i/\partial n_{ij}{}^2 = -\beta, \qquad \text{(A.4)}$$

$$\partial^2 g_i/\partial n_{kj}{}^2 = -\beta,\ k \neq i, \qquad \text{(A.5)}$$

$$\partial^2 g_i/\partial n_{ij}\partial n_{kj} = \partial^2 g_i/\partial n_{kj}\partial n_{ij} = \gamma\prod_{s\neq i,k} n_{sj},\ k \neq i. \qquad \text{(A.6)}$$

(A.2) and (A.3) imply that for $n_{ij} = n_{kj}$ for all $k \neq i$,

$$\partial g_i/\partial n_{ij} - \partial g_i/\partial n_{kj} = \alpha(\theta^i{}_i - \theta^i{}_k) > 0, \qquad \text{(A.7)}$$

as required by Assumption 1, and that for any $(n_{1j},n_{2j},\ldots,n_{Ij}) > 0$,

$$\partial g_i/\partial n_{ij} - \partial g_k/\partial n_{ij} = \alpha(\theta^i{}_i - \theta^k{}_i) > 0, \qquad \text{(A.8)}$$

as required by Assumption 2. For $I = 2$, $g_i(\cdot)$ in (A.1) is strictly concave for $\beta > \gamma$. However, for $I > 2$, the conditions for strict concavity depend on the n_{ij}. For example, for $I = 3$, strict concavity requires $\beta^2 - \gamma^2 n_{kj}{}^2 > 0$, $k = 1,2,3$, and $-\beta^3 + \beta\gamma^2(n_{1j}{}^2 + n_{2j}{}^2 + n_{3j}{}^2) + 2\gamma^3 n_{1j}n_{2j}n_{3j} < 0$.

Using the form for $g_i(\cdot)$ from (A.1) above and the linear form for $c(\cdot)$ from (III.7), the first order conditions for a type-i optimal cluster become:

$$\partial V_i/\partial n_{ij} = \partial g_i/\partial n_{ij} - t = \theta^i{}_i\alpha - \beta n_{ij} + \gamma\prod_{s\neq i} n_{sj} - t = 0 \qquad \text{(A.9)}$$

$$\partial V_i/\partial n_{kj} = \partial g_i/\partial n_{kj} - t = \theta^i{}_k\alpha - \beta n_{kj} + \gamma\prod_{s\neq k} n_{sj} - t = 0 \qquad \text{(A.10)}$$

$$\partial V_i/\partial n_{k'j} = \partial g_i/\partial n_{k'j} - t = \theta^i{}_{k'}\alpha - \beta n_{k'j} + \gamma\prod_{s\neq k'} n_{sj} - t = 0 \qquad \text{(A.11)}$$

Subtracting (A.10) from (A.9) gives, after a bit of algebra,

$$n^{*i}{}_{ij} - n^{*i}{}_{kj} = \alpha(\theta^i{}_i - \theta^i{}_k)/(\beta + \gamma\prod_{s\neq i,k} n_{sj}) > 0,\ k \neq i, \qquad \text{(A.12)}$$

while subtracting (A.11) from (A.10) gives

$$n^{*i}_{kj} - n^{*i}_{k'j} = \alpha(\theta^i_k - \theta^i_{k'})/(\beta + \gamma\prod_{s\neq k,k'} n_{sj}),\ k' \neq k. \quad \text{(A.13)}$$

(A.12) shows that since $\theta^i_i > \theta^i_k$ for all $k \neq i$, $n^{*i}_{ij} > n^{*i}_{kj}$. This is a specific instance of the plurality result from Proposition 1. (A.13) shows that n^{*i}_{kj} is greater than, equal to, or less than n^{*i}_{kj} as θ^i_k is greater than, equal to, or less than $\theta^i_{k'}$, $k' \neq k$. In this example, a higher value of θ^i_k corresponds to a higher marginal benefit $\partial g_i/\partial n_{kj}$. Thus, (A.13) indicates that there will be more type-k workers than type-k′ workers in a type-i optimal cluster if type-k workers have the larger marginal impact on the productivity of type-i workers.

For I = 2, the first order conditions that characterize a type-1 optimal cluster are

$$\partial V_1/\partial n_{1j} = \theta^1{}_1\alpha - \beta n_{1j} + \gamma n_{2j} - t = 0, \quad \text{(A.14)}$$

$$\partial V_1/\partial n_{2j} = \theta^1{}_2\alpha - \beta n_{2j} + \gamma n_{1j} - t = 0, \quad \text{(A.15)}$$

while the first order conditions that characterize a type-2 optimal cluster are

$$\partial V_2/\partial n_{1j} = \theta^2{}_1\alpha - \beta n_{1j} + \gamma n_{2j} - t = 0, \quad \text{(A.16)}$$

$$\partial V_2/\partial n_{2j} = \theta^2{}_2\alpha - \beta n_{2j} + \gamma n_{1j} - t = 0. \quad \text{(A.17)}$$

Solving (A.14) and (A.15) gives

$$(n^{*1}{}_{1j}, n^{*1}{}_{2j}) = (1/(\beta^2 - \gamma^2))(\alpha(\beta\theta^1{}_1 + \gamma\theta^1{}_2) - t(\beta + \gamma), \alpha(\gamma\theta^1{}_1 + \beta\theta^1{}_2) - t(\beta + \gamma)). \quad \text{(A.18)}$$

Solving (A.16) and (A.17) gives

$$(n^{*2}{}_{1j}, n^{*2}{}_{2j}) = (1/(\beta^2 - \gamma^2))(\alpha(\beta\theta^2{}_1 + \gamma\theta^2{}_2) - t(\beta + \gamma), \alpha(\gamma\theta^2{}_1 + \beta\theta^2{}_2) - t(\beta + \gamma)), \quad \text{(A.19)}$$

where $\beta^2 - \gamma^2 > 0$ by the second order conditions. Given our earlier assumption that $\theta^i_k > 1$ for all i, $k \in \{1,2,\ldots,I\}$, $\alpha > t$ is sufficient to guarantee interior solutions.

In the I = 2 case, the contract curve $(n^{PE}{}_{1j}, n^{PE}{}_{2j})$, is given by

$$n^{PE}{}_{1j} = (1/(\beta^2 - \gamma^2))(\alpha(\beta(\lambda\theta^1{}_1 + (1 - \lambda)\theta^2{}_1) + \gamma(\lambda\theta^1{}_2 + (1 - \lambda)\theta^2{}_2) - t(\beta + \gamma)), \quad \text{(A.20)}$$

$$n^{PE}{}_{2j} = (1/(\beta^2 - \gamma^2))(\alpha(\gamma(\lambda\theta^1{}_1 + (1 - \lambda)\theta^2{}_1) + \beta(\lambda\theta^1{}_2 + (1 - \lambda)\theta^2{}_2) - t(\beta + \gamma)). \quad \text{(A.21)}$$

One can also solve for the type-optimal loci explicitly. (A.14) and (A.15) give

$$n^{*1}{}_1(n_2) = (\theta^1{}_1\alpha - t)/\beta + (\gamma/\beta)n_2, \quad \text{(A.22)}$$

$$n^{*1}{}_2(n_1) = (\theta^1{}_2\alpha - t)/\beta + (\gamma/\beta)n_1. \quad \text{(A.23)}$$

The intersection of $n^{*1}{}_1(n_2)$ and $n^{*1}{}_2(n_1)$ gives the type-1 optimal city in (A.18). Similarly, (A.16) and (A.17) give

$$n^{*2}{}_1(n_2) = (\theta^2{}_1\alpha - t)/\beta + (\gamma/\beta)n_2, \quad \text{(A.24)}$$

$$n^{*2}{}_2(n_1) = (\theta^2{}_2\alpha - t)/\beta + (\gamma/\beta)n_1, \quad \text{(A.25)}$$

and their intersection gives the type-2 optimal city in (A.19).