Book review

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

This book explores the economic reasons for and consequence of firms and households cluster in regions, cities, and commercial districts. It provides comprehensive characterizations of the patterns and scope of these clusters when generated under alternative market structures and evaluates the efficiency and equity implications of the resulting allocations. Adopting Papageorgiou's [10] view that it would be futile to look for the model explaining different types of economic agglomerations, the book uses large variety of sophisticated models to explore agglomerations in different "geographical scales" and "degrees of sectoral details". A substantial part of the analysis is based on earlier works by the authors and other relevant literature. In particular, the book is closely related to that of Fujita et al. [2]. However, the main concerns of the two books and their framework of analysis differ considerably. Fujita et al. [2] endeavors to integrate urban economics, regional science, and international trade in a unified framework based on the emergence of agglomeration under monopolistic competition due to a preference for varieties and scale economies on the firm's level. The book under review is mainly concerned with the phenomenon of agglomeration per se with a special focus on cities. Furthermore, it adopts a much wider perspective that allows it to explore the reasons for the emergence of agglomeration under alternative market structures. In its quest to provide comprehensive treatment of the issue, the authors rely heavily on the previous studies on agglomeration in the urban economics and industrial organization literature than Fujita et al. [2]. Not less importantly, this book is concerned with the normative aspects of agglomeration, focusing on market failure and the distributive implications of the agglomeration process and its association with economic growth. These issues were completely ignored by Fujita et al. [2], who relegated them to subsequent studies.

The book is organized into four parts. The first part, which is subdivided into five chapters, deals with the fundamentals of geographical economics. Chapter 2 mainly concerns the failure of the (perfect) competitive paradigm to explain the emergence of specialization, trade and, consequently, agglomeration when space is homogenous (Starrett's "spatial impossibility theorem"). It also discusses the two welfare theorems when equilibrium is viable (e.g., when the landscape is heterogeneous).

Chapter 3 deals with von Thünen's agricultural model and its application to the monocentric urban form paradigm. The imposition of a city which serves as a market or a CBD violates the premise of the spatial impossibility theorem that the landscape is homogeneous and, thus, allows the emergence of an efficient competitive market allocation.

Chapter 4 concentrates on the fundamental trade-off between scale economies and transportation in generating agglomeration and determining its scope. When communities can capitalize the land rent into their payoff, they can offset the losses associated with scale economies and marginal cost pricing (the Henry George Rule). Then competition among profit-maximizing entrepreneurs yields an optimal allocation. Two alternative micro-economic foundations are used to explain the reduced form representation of scale economies in the city's aggregate production function. The first is the advantage of large varieties of intermediate goods which are accessible in large cities. The second is the advantage of large cities in allowing a better match of heterogenous workers and firms' jobs which leads to higher wages. Higher wages compensate the workers for the higher cost of living in large cities, caused by the higher locational costs (commuting and rent). This chapter also discusses spatial competition and how the introduction of land market may help to eliminate the markup and the excessive number of firms.[1]

The first part of Chapter 5 differs from Chapter 4 in that the scale economies inherent in the provision of a pure local public good replaces the scale economies in the aggregate production function. The chapter also elaborates on the determination of number and size of public facilities under politics. City residents vote on the number of facilities and their location (in two steps) and the facilities are financed by a proportional income tax. Without land market, the voting outcome maximizes the Rawlsian but not the Benthamite social welfare function. With land market and free mobility, individual's utility is equalized across locations and the equalized utility is maximized.

The second part of this volume, which is subdivided into two chapters, is concerned with the structure of a metropolitan areas. Using the findings of a series of earlier studies, which are based on Solow and Vickrey's [11] modelling of communication interactions, the chapter explains the emergence of symmetrical unimodal rent and density functions in urban areas, both in residential cities, when the interactions involve individuals and in the CBD when the interactions involve firms. The chapter continues by examining the effect of two types of interactions: communication interactions among firms and interactions between individuals and firms (employment). The fundamental dilemma is the attraction of firms to each other in order to reduce the impediment of distance on the marginal productivity of communication, on the one hand, and the attraction of employees' housing to firm's location to save on commuting costs, on the other. In determining its location, each firm maximizes profit given the locational variation of land rent, wage, and the aggregate net advantage to productivity of communication with other firms. In determining residential location, the individual maximizes utility given the locational variation of wage and land rent. Depending on the functional specifications and

---

[1] The introduction of land market, however, does not allow full decentralization through price-taking competition; an appropriate institution (a planner) is still required to allow each entrant to fully capitalize the increment in aggregate land rent given the predetermined expenditure of the consumers.

parameters, the model yields alternative land use patterns, including monocentric, incompletely integrated land use, completely integrated land use, and several variations of polycentric cities. Independently of the specification, however, the model implies one common result: given the parameter representing the impediment of distance on the marginal productivity of communication interactions, a monocentric configuration emerges when commuting costs are sufficiently low and full integration emerges when they are sufficiently high. On a normative level, due to the external effect of communication (the agent ignores the advantage of its location decision on the benefit of communication interactions accruing to other agents), the spatial equilibrium tends to be insufficiently agglomerated.

Chapter 7 discusses clustering of firms within cities under monopolistic and oligopolistic competitions, illustrating that imperfectly competitive markets are a major reasons for the existence of agglomeration. The first model describes how alternative residential and commercial land use patterns are generated under monopolistic competition when the representative consumer's utility (entropy-type specification) exhibits preference for varieties. In maximizing its objective function, each type of agent is spatially attracted to each agent of other type, whereas each agent competes on land with other agents, whether belonging to his own type or not.[2] There are two basic equilibrium land use patterns, mixed land use surrounded by either exclusive residential or exclusive business land use. The one that prevails depends on the average number of customers per store. In each case, the equilibrium and the first-best land use patterns are the same.[3]

The second model is an extension of the Hotelling [5] spatial competition problem. Firms compete on price and location, where the utility of a representative consumer exhibits preference for varieties (entropy-type specification). A sufficient product differentiation and low transport costs yield clustering of all the firms in the market center under Nash equilibrium. The reason being that product differentiation relaxes price competition in the cluster and low transport costs implies that only a small segmented market can be secured by moving away from the cluster. This concentration is optimal and, with free entry of firms, their equilibrium number is also close to optimum.[4] Location concentration under Nash equilibrium may, however, be higher than first-best optimum, if transport costs are sufficiently high and or product differentiation is sufficiently low.

The third main issue, studied in Section 7.4, is clustering of shops when heterogeneous customers do not know the variety each store offers and, therefore have to travel for searching among stores before choosing where and what to buy. Formally, consumers are distinguished by two parameters: location in a linear geographical space and idiosyncratic ideal variety in a circular characteristics space. The consumers are exogenously distributed on the envelope of a cylinder that portrays the geographical-characteristics space. The issue is to characterize the firms distribution in the geographical space under Nash equilibrium when only the firms' locations is known and the range of varieties offered in

---

[2] This description deviates somewhat from verbal description in the book.
[3] This does not necessarily imply that the Nash equilibrium land use pattern is efficient because, given the markup pricing, efficiency requires a second best land use pattern, not a first-best.
[4] Once again, it is not necessarily efficient because, given the markup, a second-best allocation is required.

the market, but not the variety offered by each firm. Lack of information about the address where each variety is offered and rational search strategy is shown to yield clusters of stores which need not be at geographical center.

The formation of employment centers in a land use model is the subject of Section 7.5. There is a continuum of identical exporting firms, each selling its output on the world market at a fixed price. Each firm uses a fixed amount of labor and land and variable quantities of differentiated intermediate goods. The contribution of each intermediate good to the output of an exporting firm is represented by an entropy-type function of its quantity. It follows that the exporting firm's output increases with the mass of the intermediate goods. This model yields similar results to that of the residential–business competition model in Chapter 6 when we substitute exporting firms and specialized firms for housing and firms, respectively. Thus, integrating the two models allows higher resolution of urban mixed land use and a richer gallery of land use patterns than suggested by each model separately.

The third part, which is subdivided into two chapters, discusses inter-regional industrial location. Chapter 8 discusses three models that portray industrial agglomeration under Marshallian externalities. According to the first, the centripetal force is generated by the positive scale effect of mobile skilled workers and the centrifugal force is created by the negative crowding effect of the aggregate local labor force, which includes both the mobile skilled and immobile unskilled workers. Each firm's production function is a product of a CRS function of skilled and unskilled workers and a positive external effect of the aggregate skilled workers in the region. Each firm takes the external effect as given and, therefore, considers its production function to exhibit a CRS production function. Hence, the allocation is determined competitively. Skilled workers are attracted to that region which offers them higher utility. Stability implies that if the external effect of skilled workers is sufficiently strong, the skilled workers are (partially) concentrated in one region, and unskilled workers in that region earn a higher wages than unskilled workers in the other region. Otherwise, only a symmetrical allocation is a stable equilibrium. The second and the third models discuss cases where immobile population with quadratic utility is distributed equally between two regions. The analysis in both models focuses on the effect of localization economies on the firm's location decision. In the second model, it is assumed that the advantage of localization economies varies between the regions and two firms compete strategically in price and location. The first-best allocation implies that if transport costs are sufficiently low, the two firms should be located in the region where the localization economies are larger, whereas strategic competition leads to such an agglomeration only if the localization economies in that region are sufficiently large. The third model examines the effect of localization economies with a continuum of firms each of which produces differentiated goods. The model implies that when transportation costs decline, the allocation is transformed from perfect symmetry to agglomeration. For sufficiently low transport costs, all the firms agglomerate in one region. The first-best allocation is never less agglomerated than the market equilibrium.

Chapter 9 is devoted to the "new economic geography" where agglomeration in the form of core–periphery structure is generated under monopolistic competition by preference for varieties and scale economies on a firm level. Beginning with the original version of the core–periphery model where the subutility of differentiated products is a

CES function and the transport costs are represented by melting iceberg, it is shown that agglomeration, represented by a core–periphery structure, requires low transport costs. This observation is also corroborated by a modified model in which the subutility of the differentiated products is quadratic and the transport cost of a unit shipped between the regions is fixed in terms of the numeraire. In contrast to the original version, however, the modified version of the core–periphery model lend itself to welfare analysis. It is shown that agglomeration helps the mobile skilled workers who concentrate in the core and the immobile unskilled workers there, but hurts the immobile unskilled workers in the periphery. The loss of the unskilled workers in the periphery may outweigh the gain of the unskilled workers in the core. The second version of the core–periphery model is extended to include a mobile intermediate sector and perfect foresight.

The fourth part of the book, which is subdivided into two chapters, is concerned with urban system and regional growth. Chapter 10 discusses the stability of von Thünen's monocentric structure of and the evolution of the urban system when the population grows. The stability of the monocentric structure depends on low transport costs of the agricultural product relative to the cost of moving manufactured goods. When the population increases, the demand for agricultural products increases and the agricultural hinterland expands. The local prices of differentiated goods increase in the periphery, which attract manufacturing firms and, hence, new cities are established. By adding an intermediate goods sector to the model, a rich variety of results can be derived such as the emergence of a regular pattern of cities in accordance with some characteristics of the central place theory. Yet, the model does not yield a perfect hierarchy where each city trades with a limited set of cities and its agricultural hinterland. Rather, each city, which, according to the present model, specializes in its unique set of differentiated goods, trades with each of the other cities.

The volume culminates in Chapter 11, which synthesizes the core–periphery model and growth theories and evaluates the prevailing policy aimed at restricting agglomeration. The basic idea is that knowledge is accumulated at a rate equal to the aggregate production of patents by skilled workers. It is assumed that the productivity of patents production is (indirectly) proportional to the stock of knowledge, where the coefficient of this proportion in a given region increases with its share in the aggregate skilled workers (reflecting the decay of personal knowledge spillover with distance). This assumption underlies the crucial association between the rate of increase of knowledge (aggregate output of patents) and agglomeration. Each new patent allows the establishment of a new firm which produces a new brand of differentiated goods. Hence, the rate of aggregate increase of varieties, which is a measure of the rate of growth, is increased with agglomeration. Solving the steady state growth path, it is shown that, with perfectly mobile patents, the R&D sector as well as a substantial part of manufacturing firms agglomerate in one region and when the transport cost is sufficiently low, both the R&D sector and the entire manufacturing sector agglomerate in one region. Under a core–periphery steady-state growth path, the welfare of unskilled workers in the core exceeds that of unskilled workers in the periphery. The welfare of the latter may, but need not, be hurt relative to what they enjoy under a symmetric growth path. Yet, a dispersion policy is likely to reduce global economic growth and may even hurt the unskilled workers in the periphery. This policy, however, may be justified when the criterion of horizontal equity is adopted and there is no way to compensate the workers in the periphery.

This volume provides the most comprehensive study of the fundamental reasons for agglomeration and the market incentives that drive agents to agglomerate. It is distinguished from earlier studies by its special focus on intra-urban agglomeration, integration of agglomeration analysis in urban and regional economics as well as industrial organization literature, novel exploration into normative issues associated with agglomeration and growth, and last but not least, integration of agglomeration with growth models. No student of agglomeration can do without this study. Yet, I think that a coherent epilogue (Chapter 12) is missing in this book. Indeed, at the book's outset (page 2), the reader is "warned" not to expect to find the theory of agglomeration. Rather, citing Papageorgiou's [10] conjecture that each type of agglomeration requires a different model, the authors offer a rich gallery of models, each adapted to different scale and sector. Accordingly, the quest to generalize results is confined to the concluding comments of the relevant chapters with only occasional cross reference to other similar models or models that yields opposing results. I do think that even in the case portrayed by Papageorgiou [10], a coherent epilogue is required for presenting the most concise set of explanations for the emergence of agglomeration, categorizing them, and surveying the main results derived in the book, while spelling out what is common and what is not and why. I will elaborate on this in the sequel.

First, consider the following list of cited statements regarding the reasons for agglomeration (it is not an exhaustive list):

1. "According to Marshall (1890, 1920, Chap. X), externalities are crucial in the formation of economic agglomerations...", (p. 7).
2. "The hypothesis that production sets are convex implies that production exhibits no increasing returns to scale—whatever is scale... If the distribution of natural resources is uniform, the economy is such that each person produces for her own consumption, we therefore have backyard capitalism." (p. 27).
3. "1. ... the advantages that firms producing similar goods may exploit by collocating. For example, different agents own different bits of information and their gathering yields higher level of knowledge... 2. A large market... allows for a large of intermediate commodities... intermediate commodities can be used as inputs to enhance the productivity of the final sector... 3. a large city allows for a better average match between heterogeneous workers and firms' job requirements." (p. 98).
4. "But why do households and firms seek proximity? Fundamentally, this occurs because economic agents need to interact and distance is an impediment to interaction... one can view the agglomeration process... as an interplay between an interaction field among agents and competition on land market... the need to interact acts as a centripetal force, whereas competition for land has the nature of centrifugal force." (p. 169).
5. "...externalities are at the root of economic agglomerations." (p. 210).
6. "Such an externality effect may explain why economic agents are prepared to pay high rent to live close to the centers of large cities where this effect is more intense" (p. 211).
7. "...the presence of imperfectly competitive markets is another major reason for the existence of...agglomerations." (p. 258).

I would like to evaluate these statements in the light of what I learned from reading the book on the reasons for agglomeration. It can be summarized as follows: Convexity of the production set is indispensable for price-taking equilibrium. However, such convexity implies that, under the competitive paradigm, firms are motivated to disperse production according to the spatial distribution of demand in order to save on transport costs. Hence, in order to explain clustering of firms in relatively small areas, we need either non-convexity of the production set and, thus, non-competitive market structures or, alternatively, some centripetal forces which pull together either individuals (as consumers and workers) to individuals, firms to firms, and clusters of individuals to clusters of firms. Non-convexity of collective goods production and better matching of individuals in their social activities in large groups may explain the centripetal force which pulls individuals together. The advantage of information exchange may explain the centripetal force which pulls firms together. A better average match between heterogeneous workers and firms' job requirements (on a regional scale) and commuting costs (on a city scale) underlie the spatial attraction of individuals as employees to firms. An additional complication is added when the same force is both centrifugal and centripetal. Consider the effect of commuting costs on metropolitan structure (see discussion of Chapter 6). On the one hand, they induce employees to cluster close to their employment centers, thus increasing the overall residential densities. On the other hand, they attract firms to their employees to save on wages, thus reducing the density of the employment center. When commuting cost are sufficiently high, the city core, which accommodates the employment and shopping centers, may be dismantled.

According to the above approach, statements 2, 4 and 7 in the above list explain agglomeration, 3 includes special cases of 4, but statements 1, 5, and 6 are questionable. Indeed, pursuing the advantage of interactions, individuals and firms make some decisions which have external effects. For example, consider Chapter 6, where, in making a location decision, individuals and firms ignore the effect of their location decision on the benefit of communication interchange derived by other agents. The result is insufficient agglomeration.

Suppose, for the sake of argument, that firms are altruists and, in making their location decision, take into consideration the communication benefit to other firms, such that the external effect disappears. Would not then the firm be willing to pay higher rent for locating closer to the city center than otherwise? Will not the agglomeration prevail and even increase? Furthermore, restricting ourselves to selfish firms, suppose that the city is run by a developer who maximizes the surplus of the city (the aggregate output minus the minimum aggregate cost required to allow the residents to achieve their reservation utility). Will not then the developer agglomerate firms in order to increase their productivity? Will not the resulting agglomeration be even higher than with the externality that distorts the allocation? All this boils down to my contention that externality deters rather than enhances agglomeration, as implied by the cited statements 1, 5, and 6 above. According to this approach, the fundamental advantage of interactions between agents mentioned in cited statement 4 is unduly referred to as "externality" and citing Marshall in this context is not a vindication.[5]

---

[5] The term "externality" is vague in the economic jargon (see Lagueux [7]). One issue is whether externality can be identified by information on the fundamentals alone. The answer of Arrow [1] is negative.

   The book could have also been benefited from including in the proposed epilogue a coherent discussion of the main results derived by each model and explaining the reasons for their differences. Two issues which occupy the discussion of most of the models are the effect of transport cost on agglomeration and the market bias in inducing agglomeration vis à vis optimum.

   Beginning with the first issue, the standard core–periphery model implies that agglomeration requires low transport costs of manufactured goods. The same result is obtained when a quadratic utility and a linear transport cost function are, respectively, substituted for the CES utility and melting iceberg transport functions, as well as in some other models (see the models portraying industrial agglomeration under Marshallian "externalities" under duopoly and Chamberlin's large group competition in Sections 8.3 and 8.4, respectively). Yet, when the source of the centrifugal force is changed from immobility of unskilled workers to urban land scarcity, as in Helpman [4], Tabuchi [12], and Ottaviano et al. [9], the above functional relationship between transport cost and agglomeration collapses. Tabuchi [12] and Ottaviano et al. [9] derive an inverted u-shaped relationship, whereas Helpman [4] shows that agglomeration requires sufficiently high transport costs rather than sufficiently low. Furthermore, when intermediate goods are introduced into the core–periphery setup, the relationship between agglomeration and transport cost turns to be three dimensional. A core–periphery structure still requires that the costs transporting the manufactured good should be sufficiently low. However, the transport costs of the intermediate goods should be sufficiently high. Krugman and Venables [6] assume that the same varieties are both final and intermediate goods and derive an inverted u-shaped relationship, similar to Tabuchi [12] and Ottaviano et al. [9]. To make things even more complicated, in the von Thünen setup (without intermediate goods), transporting manufactured goods should be sufficiently expensive relative to the cost of shipping agricultural goods. Hence, not only that each type of agglomeration requires a different model, but the same parameter may have a different qualitative effect in the margin, depending on its size.[6]

   Turning to the normative issue, the results derived and the following policy implications are not less contradicting. Most of the models discussed in the book imply that, if the market is biased, the bias is against agglomeration. This is the case with intra-city location pattern of households and firms portrayed in Chapter 6, where the reason is the presence of negative external effect. Similar externalities are present in other cases such that the market is biased against agglomeration (the markets, portrayed in Chapter 8, both in duopoly and Chamberlin's large group case, tend to induce less agglomeration than optimum). Other reasons create such a biased in the core–periphery version of Helpman [4].[7] Yet, we know from Hotelling [5] that locational competition tends to induce excessive agglomeration of firms selling identical products. Furthermore, in the oligopolistic market of retailers who sell differentiated goods, discussed in Section 7.3, optimum implies full dispersion when transport costs are sufficiently high, but "…the market may well provide insufficient geographical dispersion." This and more, in contrast

---

[6] See also Hadar and Pines' [3] evaluation of the effect of aggregate population size on agglomeration in a two region model.
[7] This result is not referred to in the book.

to the case of Helpman [4], in the linear model of core–periphery (Section 9.4), the market may (for intermediate values of transport costs) provide excessive agglomeration. The contradicting implications, especially in the core–periphery models, do not allow us even today to evaluate the contention that "...the largest urban areas are too large and the smallest too small.", which Mills and Hamilton (p. 366) [8] considered to be "more widely held than any other regarding social problem".

References

Arrow, K.J., 1970. Political and Economic Evaluation of Social Effects and Externalities. In: Margolis, J. (Ed.), The Analysis of Public Output. Columbia University Press, New York, pp. 1–23.

Fujita, M., Krugman, P., Venables, A.J., 1999. The Spatial Economy: Cities, Regions, and International Trade. MIT Press, Cambridge, MA, pp. 1–23.

Hadar, Y., Pines, D., 2004. Population growth and its distribution between cities: positive and normative aspects. Regional Science and Urban Economics 34, 125–154.

Helpman, E., 1998. The size of regions. In: Pines, D., Sadka, E., Zilcha, I. (Eds.), Topics in Public Economics. Cambridge University Press, New York, pp. 33–54.

Hotelling, H., 1929. Stability in Competition. Economic Journal 39, 41–57.

Krugman, P., Vanables, A.J., 1995. Globalization and the inequalities nations. Quarterly Journal of Economics 110, 857–880.

Lagueux, M., 1998. Learning from the debate on externalities. In: Backhouse, R.E., Houseman, D.M., Maki, U., Salanti, A. (Eds.), Economics and Methodology. MacMillan, London, pp. 120–147.

Mills, E.S., Hamilton, B.W., 1989. Urban Economics. Harper Collins, New York.

Ottaviano, G.I.P., Tabuchi, T., Thisse, J.-F., 2002. Agglomeration and trade revisited. International Economic Review 43, 409–436.

Papageorgiou, Y.Y., 1983. Models of agglomeration. Sistemi Urbani 3, 391–410.

Solow, R., Vickrey, W.S., 1971. Land use in a Long Narrow City. Journal of Economic Theory 3, 430–447.

Tabuchi, T., 1998. Urban agglomeration and dispersion: a synthesis of Alonso and Krugman. Journal of Urban Economics 44, 333–351.

D. Pines
The Eitan Berglas School of Economics,
Tel Aviv University, Ramat Aviv, 69978, Tel Aviv, Israel
Email-address: pines@post.tau.ac.il.
Tel.: +972 3 6409904; fax: +972 3 6409908.