ABSTRACT

In recent years, monopolistic competition models have frequently been applied in macroeconomics, international and interregional economics, and economic growth and development. In this paper, I present a highly selective review in this area, with special emphasis on the complementarity and its role of generating multiplier processes, agglomeration, underdevelopment traps, regional disparities, and sustainable growth, or more generally, what Myrdal (1957) called the "principle of circular and cumulative causation."

*Much of my own work reported here has been done while I was visiting the Hoover Institution at Stanford University, the Federal Reserve Bank of Minneapolis, and the Institute for International Economic Studies, University of Stockholm, Sweden. I would also like to thank the Bank of Japan for giving me the opportunity to report this project. Any views expressed here are mine and should not be attributed to any other individual or organization. Any opinions, findings, conclusions, or recommendations expressed herein are those of the author and not necessarily those of the National Science Foundation, the University of Minnesota, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.
1. Introduction

In recent years, the paradigm of monopolistic competition has been applied to model a variety of aggregate phenomena, particularly in macroeconomics, international and interregional economics, and economic growth and development. Monopolistic competition can be defined as a situation of imperfect competition with the following features:

a) The products sold are differentiated. Each firm, as a sole producer of its own brand, is aware of its monopoly power and sets the price of its product.

b) The number of firms (and products) is so large that each firm ignores its strategic interactions with other firms; its action is negligible in the aggregate economy.

c) Entry is unrestricted and takes place until the profits of incumbent firms are driven down to zero.

These features make the paradigm of monopolistic competition very useful for modelling aggregate phenomena. First of all, as a form of imperfect competition, it allows us to describe decentralized allocations in the presence of increasing returns. Second, unlike oligopoly models, it helps us to focus on the aggregate implications of increasing returns without worrying about strategic interactions among firms and the validity of profit maximization as the objective of firms. Third, in models of monopolistic competition, the range of products supplied in the market is endogenously determined through entries and exits. This feature makes monopolistic competition a useful apparatus within which to formalize growth and development processes, as economics grow and our standards of living rise not so much by producing or consuming more of the same products, but by adding new products to the list of those we already produce and consume.

In this paper, I intend to present a highly selective review of recent developments in this area. Central to the following discussion is the notion of complementarity. Broadly speaking, complementarities are said to exist when two phenomena (or two actions, or two activities) reinforce with each other. For example, if an expansion of industry A leads to an expansion of industry B,
which in turn leads to an further expansion of A, then the two industries are complementary to each other. Or if the arrival of a new store into a town makes it a desirable location for other stores, then there are complementarities in the locational decisions. Such a complementarity introduces some circularity in the economic system, which has profound implications on the stability of the system. If a change in certain activity is initiated by an exogenous shock, this leads to a similar change in complementary activities and starts a cumulative process of mutual interaction in which the change in one activity would continuously be supported by the reaction of the others in a circular manner. Many writers in the past, such as Hicks (1950), Kaldor (1985), Kalecki (1939), Myrdal (1957), and Nurkse (1953), among others, have stressed that the cumulative processes of this kind should be an essential element in explaining business cycles, under-development, economic growth, and regional inequalities.

On the other hand, the standard neoclassical paradigm, exemplified by Arrow and Hahn (1971), emphasizes self-adjusting mechanisms of the market forces. The stability of equilibrium in the standard theory comes from resource constraints. As different activities compete for scarce resources, an expansion of one activity comes only at the expense of others. In the system with well-functioning markets, the resource allocation is efficient, which leaves little room for complementarity and circular causation. Imperfect competition and increasing returns, as departures from the standard paradigm, make the system less stable and more conducive to complementarities and cumulative processes.

Ever since the publication of the famous book by Edward Chamberlin in 1933, a large number of studies has applied the paradigm of monopolistic competition in a variety of contexts. Indeed, the literature already has many excellent surveys, such as Beath and Katsoulacos (1991), Benassy (1991), Eaton and Lipsey (1989), Stiglitz (1986), and Tirole (1988). These studies, however, are written in a different perspective and scope. They address the questions like; how the models of monopolistic
competition differ from oligopoly models, or what is a most appropriate way of modelling product differentiation, or whether the market equilibrium provides the optimal product diversity and selection. In other words, they are written predominantly in partial equilibrium or industrial organization perspectives. Some studies, such as Benassy (1991) and Hart (1985), discuss general equilibrium models of monopolistic competition, but their concerns remain largely theoretical, such as the existence of equilibrium, its uniqueness, and the limit theorem. In this paper, on the other hand, I will present how general equilibrium models of monopolistic competition can be applied to explain complementarities and cumulative phenomena and discuss their implications in the context of macroeconomics, international and regional economics, as well as growth and development.

2. The Monopoly Pricing Distortion and Multiplier Processes

The departure from perfect competition means that the firm, faced with downward sloping demand, sets the price above the marginal cost. In the presence of such a distortion, aggregate demand management could be effective in stimulating the aggregate economic activity, as well as in raising the welfare of the economy. To grasp the intuition behind the mechanism, suppose the government increases its demand for monopolistically competitive goods, financed by a lump-sum tax. Since the prices exceed the marginal costs, such a shift in demand increases the level of monopoly profits in the economy and thus the national income. This increased income will generate additional demand for monopolistically competitive goods, which further raises profits and income and so on. With the monopoly price distortion, the equilibrium behavior of the economy resembles very much like the multiplier process described in a simple textbook Keynesian model.

The Simple Model

The idea can be modelled as follows. To focus on the monopoly pricing distortions, I will
ignore the entry process and endogenous changes in the product variety. In other words, it is assumed that the economy produces a fixed set of products, each of which is supplied by a sole monopolist. Let \( z \in [0,1] \) be an index of a product, as well as the monopolist firm producing it. The assumption of restricted entry may be defended by asserting that some specific factors, such as entrepreneurial capital, are required to operate firms, and they are fixed in supply. Or one could simply argue that the model describes the short-run equilibrium.

The representative consumer is endowed with \( L \) units of labor, holds the ownership share of the profit-making firms, and maximizes the following preferences,

\[
\alpha \int_0^1 c(z) \, dz + (1-\alpha) \ln(N),
\]

where \( c(z) \) denotes consumption of variety \( z \), and \( N \) leisure, respectively; \( \alpha \) represents the marginal budget share of the product group, \( z \in [0,1] \), and is assumed to be between zero and one. Taking the leisure as a numeraire, the budget constraint is given by

\[
\int_0^1 p(z) c(z) \, dz + N \leq Y - T,
\]

where \( Y = L + \Pi \), represents the aggregate income, equal to the sum of the labor income, \( L \), and the profit, \( \Pi \), while \( T \) is the lump sum tax. As a solution to this consumer's problem, one can obtain the demand for each variety of goods as,

\[
c(z) = \frac{\alpha(Y - T)}{p(z)},
\]

Let us assume that the government spends \( G \) on each good, equally across all goods regardless of their prices. Then, the total demand for each good is given by \( q(z) = \{\alpha(Y - T) + G\}/p(z) \). The government also hires \( N' \) units of labor; the budget constraint requires \( T = G + N' \).

Each variety can be produced by two types of firms. First, there is a competitive fringe of
firms that convert one unit of labor input into one unit of output with constant returns to scale technology. Second, there is a unique monopolist firm with access to an increasing returns to scale technology. This firm alone can produce \( q \) units of output by using \( aq + F \) units of labor input, where \( 0 \leq a < 1 \), and \( F \) represents the fixed cost. This firm chooses \( p(z) \) to maximize its profit, \( \pi(z) = p(z)q(z) - [aq(z) + F] \). In doing so, it treats \( C \) and \( G \) as fixed parameters; although this firm has some monopoly power over its own variety, it is negligible relative to the aggregate economy. Because of the unit elasticity of the total demand and the competitive fringe, the monopolist practices the limit pricing, \( p(z) = 1 \), and thus \( \mu = 1 - a = \{p(z) - a\}/p(z) \) can be interpreted as the profit margin. Since all monopolists face the same incentive, \( p(z') = 1 \) and thus \( c(z) = C = \alpha(Y - T) \) and \( q(z) = \alpha(Y - T) + G \) for all \( z \in [0,1] \). The aggregate profits are therefore equal to \( \Pi = \pi(z) = \mu[\alpha(Y - T) + G] - F \). Note that higher aggregate demand increases aggregate profits in the presence of the profit margin.

The income identity implies \( Y = L + \Pi = L + \mu[\alpha(Y - T) + G] - F \), or

\[
Y = A + \mu \alpha Y, \tag{1}
\]

where

\[
A = L - F + \mu (G - \alpha T)
\]

is the "autonomous" component of the aggregate income. Solving the income identity (1) for the equilibrium income yields

\[
y = \frac{A}{1 - \mu \alpha} = \frac{L - F + \mu (G - \alpha T)}{1 - \mu \alpha}. \tag{2}
\]

Note that one unit increase in the government spending on the monopolistically competitive products raises \( A \) by \( \mu \) when unaccompanied by a tax increase, and by \( \mu(1 - a) \) when financed by the lump-sum tax. This autonomous increase in income generates an induced demand increase by \( a \), hence further
increasing the income by $\mu \alpha$. Through such a cumulative process, the aggregate income increases by the amount equal to the original increase in the "autonomous" component of aggregate income, "multiplied" by $1 + \mu \alpha + (\mu \alpha)^2 + ... = 1/(1-\mu \alpha)$.

**Aggregate Demand Spillovers**

It is easy to show that such an aggregate demand policy improves the welfare of the representative consumer. This is because in equilibrium the non-competitive goods are consumed too little, due to the monopoly pricing. Alternatively, this inefficiency can be understood in terms of aggregate demand spillovers. To see this, suppose that all consumers simultaneously increase their demand; this leads to a more efficient allocation of resources, as the marginal benefit of consumption exceeds the social cost of production. Nevertheless, any individual consumer has no incentive to demand more in equilibrium. The discrepancy between the effects of coordinated versus unilateral demand increases arise, as the potential gains generated by a unilateral shift in demand, materialized as an increase in the monopoly profit, will be widely dispersed in the economy. This spillover effect creates sort of a free-rider problem in the consumption decision. The government spending improves welfare as it solves the free-rider problem.

Benassy (1978) and Negishi (1978, 1979) earlier demonstrated underemployment and insufficient aggregate demand in monopolistically competitive economies, under the so-called "subjective demand" approach. For the "objective demand" approach, see Blanchard and Kiyotaki (1987), Hart (1982), Startz (1989), and Weitzman (1982). Some of these studies recently emphasized aggregate demand management policies as a way of solving the free-rider problems that arise from aggregate demand spillovers. Cooper and John (1988) discussed this issue in a more general context.
Complementarity

It is also important to note that an expansionary aggregate demand policy leads to a boom, in which all producers benefit simultaneously. The demand spillover creates a complementarity across all sectors of the economy. Such a complementarity would be more transparent when the model is extended to have two product groups, \( i = 1 \) and \( 2 \), with \( n_i \) being the size of group \( i \). Let \( \alpha_i, \mu_i \) and \( \Pi_i \) be the budget share, the profit margin, and the profit level of product group \( i \). Then, \( Y = L + \Pi_1 + \Pi_2 \) and, if the government spends \( G_i \) on product group \( i \), \( \Pi_i = \mu_i [\alpha_i (Y - T) + G_i] - n_i F \). The profits of the two groups thus satisfy

\[
\Pi_1 = A_1 + \mu_1 \alpha_1 (\Pi_1 + \Pi_2),
\]

\[
\Pi_2 = A_2 + \mu_2 \alpha_2 (\Pi_1 + \Pi_2),
\]

where

\[
A_1 = \mu_1 [G_1 + \alpha_1 (L - T)] - n_1 F
\]

\[
A_2 = \mu_2 [G_2 + \alpha_2 (L - T)] - n_2 F
\]

summarize the autonomous components of the profits in the two groups. The second terms in (3) and (4) represent the induced components; a unit increase in the total income generate additional demand for sector \( i \) by \( \alpha_i \), which increases its profit by \( \alpha_i \mu_i \). In Figure 1, the steeper line represents (3); its slope is equal to \((1 - \mu_1 \alpha_1) / \mu_1 \alpha_1\). An increase in \( A_1 \) shifts this line to the right. The other line represents (4), whose slope is \( \mu_2 \alpha_2 / (1 - \mu_2 \alpha_2) \); an increase in \( A_2 \) shifts it upward. The equilibrium profit in each group is given at the intersection. To solve explicitly,

\[
\begin{bmatrix}
\Pi_1 \\
\Pi_2
\end{bmatrix} = \frac{1}{1 - \mu_1 \alpha_1 - \mu_2 \alpha_2} \begin{bmatrix}
1 - \mu_2 \alpha_2 & \mu_1 \alpha_1 \\
\mu_2 \alpha_2 & 1 - \mu_1 \alpha_1
\end{bmatrix} \begin{bmatrix}
A_1 \\
A_2
\end{bmatrix}.
\]

Note that an increase in government spending on one group increases not only the profit (and the output) of that group, but also the profit and the output of the other group. Thus, an expansion of
one sector benefits the other sector; the two sectors now become complementary to each other through demand spillover effects.

**International Economics**

This implication can be carried over into international contexts. An increase in the autonomous demand for domestic products leads to a higher level of monopoly profits in the economy and thus of the national income. This increased income will generate additional demand for domestic products, leading to a multiplier process. To the extent that increased income in demand falls on foreign goods and raises aggregate profits abroad, it also creates similar chain reactions and lead to an increase in income abroad. Thus, under imperfect competition, there are positive spillover effects of country specific demand shocks: see Matsuyama (1992a) for a formal demonstration.

**Economic Development**

The profit multiplier process and aggregate spillover effects are also important in understanding some problems in economic development. Murphy, Shleifer and Vishny (1989), using models similar to one discussed above, demonstrated that aggregate demand spillovers, when strong enough, can create underdevelopment traps. In their models, the monopolist, instead of being equipped with the increasing returns to scale technology, decides whether to adopt the technology or to stay with the competitive fringe. Because of the fixed cost associated with the technology, the incentive to adopt depends on the market size. The market size, however, depends on the purchasing power of workers. This in turn depends on the extent of industrialization in the economy, that is, how widely the modern efficient method of production is adopted. Thus, the more firms use the modern technology, the more profitable become the use of the modern technology. The
complementarity of investment across industries leads to the co-existence of a good equilibrium and a bad equilibrium. In the former, all industries adopt modern technologies; in the latter, all industries stay with cottage production, which Murphy, Shleifer and Vishny interpreted as an underdevelopment trap. The model thus captures the old idea, which dates back to Rosenstein-Rodan (1943), Nurkse (1953), and Hirschman (1958), who emphasized the complementarity of modernization efforts across industries as the main obstacle to economic development. Rosenstein-Rodan (1943), in particular, used this idea to advocate for a large-scale development planning as a way of breaking away from underdevelopment traps. It should be stressed, however, that the complementarity of investment across industries does not necessarily provide the rationale for comprehensive central planning, as coordinated investment could be achieved through "the infectious influence of business psychology (Nurkse 1962, p.249)," which may be orchestrated by no more than some form of indicative planning: see Matsuyama (1992b) more on this issue.

In the models discussed above, the range of products in the economy is fixed, and hence monopoly rents are not dissipated away by the process of entry. Once unrestricted entry is allowed for, monopoly profits would disappear and so would the complementarity through the profit multiplier process. Nevertheless, free entry brings another source of complementarity as the entry of new firms generally expands the variety of products supplied in the market. I will turn to this issue for the remainder of the paper.

3. **Expanding Product Variety and Increasing Returns**

To understand how the entry process could lead to a complementarity, it is important to note that entry of new firms, by introducing new products and services to the market, gives rise to increasing returns at the aggregate level. The idea itself is not new. Young (1928), for example,
stressed that progressive division and specialization of industries, rather than subdivision of labor within a firm, as an essential part of the process by which increasing returns are realized. The formal modelling of this idea is, however, fairly recent.

**The Model**

To demonstrate the idea, I will use the following stripped down version of the Dixit and Stiglitz (1977) model. Suppose that a single consumption good is produced by assembling a variety of differentiated intermediate inputs. The technology satisfies the property of constant returns to scale for a given set of inputs. More specifically, the production function takes a form of symmetric CES;

\[ X = \left[ \int_0^n [x(z)]^{\frac{1}{\sigma}} dz \right]^{\frac{\sigma}{\sigma-1}}, \quad (\sigma > 1) \]

with \( x(z) \) being the amount of input \( z \) employed in production, and \([0,n]\) represents the range of intermediate inputs available in the marketplace. It assumes that the rate of substitution between every pair is equal to \( \sigma \). The restriction, \( \sigma > 1 \), implies that no input is essential; \( X \) is well-defined even if some differentiated inputs are not used at all. Such a restriction is necessary as we consider the situation in which the range of products offered may vary.

One implication of this specification of product differentiation deserves special emphasis. That is, the productivity of intermediate inputs increases with the range of inputs available. To see this, let us suppose that all varieties are produced by the same amount, which in fact would be the case in the optimal and equilibrium allocations. By letting \( x(z) = x \), we have

\[ X = n^{\frac{\sigma}{\sigma-1}} x. \]

Let \( M = nx \) be the total inputs used. Then, the average productivity of inputs, \( X/M = n^{1/(\sigma-1)} \), increases with \( n \). This arguably captures the notion that introduction of new capital goods and
producer services of highly specialized character would enhance the efficiency of the economy. Ethier (1982) and Romer (1987) ascribe this property of the CES specification as increasing returns due to specialization, or to the division of labor, in production.

The costs of expanding the product variety and of increasing specialization come from the economies of scale in the production of differentiated intermediate inputs. If there were no scale economies, then productivity could go up indefinitely by adding more and more varieties to the list of differentiated products, and producing less and less of each variety. As before, let us suppose that production of $x$ units of each variety requires $ax + F$ units of labor; $F$ is the fixed cost and $a$ is the marginal labor requirement. For the notational convenience, let us choose the unit of measurement so as to have $a = 1 - 1/\sigma$. Finally, the labor resource constraint is

$$L = n(\sigma_x + F), \quad (6)$$

where $L$ is the total labor supply in the economy. Combining (5) and (6) yields

$$X = \frac{1}{\sigma} \left( \frac{n}{a} \right)^{\frac{1}{\sigma-1}} (L - nF).$$

Hence, the optimal product variety, one that maximizes the above expression, and per capita consumption are

$$n^* = \frac{L}{\sigma F}, \quad \frac{X^*}{L} = \left( \frac{L}{\sigma F} \right)^{\frac{1}{\sigma-1}}.$$

They are larger a) when the fixed cost is small, b) when the products are less substitutable, and c) when the size of the economy is large.

A market equilibrium for this economy consists of the competitive final goods sector with the constant returns to scale production function $X$ and the monopolistically competitive intermediate inputs sector with the labor input function, $ax + F$. Taking the final goods price, $P$, and the prices of intermediate inputs, $p(z)$ for $z \in [0,n]$, fixed, the final goods producers seek to maximize the profit
by choosing the cost-minimizing input combination. It is straightforward to derive demand function for each input

$$\frac{x(z)}{x} = \left[ \frac{P(z)}{p} \right]^{-\sigma}. $$

Furthermore, the zero profit condition implies that the output price must be equal to the unit cost:

$$P = \left[ \int_0^x [P(z)]^{1-\sigma} dz \right]^{\frac{1}{\sigma}}. $$

Facing the demand function derived above, each intermediate producer sets the price to maximize its profit. In doing so, it takes $P$ and $X$ as fixed. This means that the elasticity of demand with respect to its own price is $\sigma$, so that the profit-maximizing price satisfies $p(z)(1 - 1/\sigma) = a$, or $p(z) = 1$. The equilibrium price of the final good hence becomes

$$P = n^{-\frac{1}{1-\sigma}}. \quad (7)$$

Note that an increasing availability of specialized inputs leads to a lower price of the output, despite each input price remains constant. This is nothing but the mirror image of the efficiency effect of increasing specialization pointed out earlier.

The gross profit of each firm (i.e., gross of the fix cost) can be shown to be proportional to the output produced:

$$\pi = (P - a) x = \frac{x}{\sigma}. \quad (8)$$

Using the labor market clearing condition (6), the gross profit can be expressed as a function of the number of firms;

$$\pi(n) = \frac{1}{\sigma - 1} \left[ \frac{L}{n} - F \right]. \quad (9)$$

Note that this is a decreasing function of $n$, as shown in Figure 2. Entry of firms hence reduces the
profit of incumbent firms. The entry process continues as long as the gross profit exceeds the fixed cost, \( F \). The unique equilibrium is thus depicted as point E in Figure 2. Some algebra shows that the equilibrium product variety and per capita consumption are

\[
n^* = \frac{L}{\sigma F}, \quad \frac{X^*}{L} = \left[ \frac{L}{\sigma F} \right]^{\frac{1}{\sigma-1}}.
\]

In this market economy, the variety of intermediate inputs, or the division of labor, is limited by the extent of the market; as the economy size increases, more firms stay and a wider range of products are offered in the market; this division of labor enhances the efficiency of production, and therefore the consumers living in a larger economy will be better off. (As the equilibrium profit is zero, the aggregate income consists only of the wage income, so per capita output is equal to the equilibrium real wage in this model.) The equilibrium output of each firm is

\[
x^* = \sigma F,
\]

which is to say that, with a large fix cost and in the presence of close substitutes, the firms need to sell more to break even.

**Economic Integration**

This model has some significant implications when applied to international and interregional economics. Imagine that there exist two economies of the kind analyzed above, say East and West, and that they are originally isolated to each other. Furthermore, assume that the two economies have identical tastes and technologies. They differ only in the labor supply; \( L \) denotes the labor supply of East and \( L^* \) that of West. Now let us ask; what would happen if they are integrated into each other? The answer to this question crucially depend on the mobility of goods and labor.

**The Effects of Trade:** First, suppose that the products can be transported at zero cost, but labor is immobile. Free trade in goods ensures that the same set of intermediate inputs are available
everywhere, so that the final goods sector in both economies achieve the same level of efficiency, while the varieties produced in each region is determined by the labor resource constraint. Symmetry implies that the number of intermediate inputs produced in each region is proportional to the labor supply; East produces \( n = L/\alpha F \) varieties, and West produces \( n^* = L^*/\alpha F \) varieties. Both regions enjoy the same level of productivity and per capita output of the final good is equal to

\[
\left[ \frac{L + L^*}{\sigma F} \right]^{\frac{1}{\sigma-1}}.
\]

As the productivity depends on the availability of differentiated inputs in this model, economic integration through trade improves the efficiency in both regions, and hence it is mutually beneficial. True, the larger economy may boast a much wider array of inputs produced than the smaller economy. However, as long as all the inputs are available, this would not handicap the smaller economy. In fact, by comparing the situations before and after the economic integration, it is easily seen that productivity gains are larger for the smaller economy.

With some notable exceptions of Negishi (1972) and others, imperfect competition and economies of scale had received little attention in the theoretical trade literature for many years. The systematic study of trade in differentiated products finally took off and has grown enormously during the last decade, following the path-breaking studies by Dixit and Norman, Krugman and Lancaster in the late seventies. I have touched on only one of many important lessons that come out of this literature: see Helpman and Krugman (1985, Chs. 6-9) for more.

The Effects of Factor Mobility: Let us now suppose that there are impediments to trade, but economic integration makes it possible for some workers to migrate across the economies. Footloose workers migrate from the smaller economy to the larger economy, where the equilibrium wage is higher. As a result, the population distribution becomes more lop-sided. New firms are created in the larger economy, while the firms are forced to close down in the smaller economy,
which makes those who stay in the smaller economy worse off. With the limited mobility of goods, increasing returns are now region-specific. This induces the footloose workers to concentrate into one region. The economies of scale are realized only through agglomeration. It should be noted that economic integration benefits the immobile workers in the larger economy, while hurting those left behind in the smaller economy. This provides a striking contrast with the case of trade in differentiated goods, where all (immobile) workers gains and, in particular, those in the smaller economy gains most. Fujita (1990, Ch. 8.4), Helpman and Krugman (1985, Chs. 10-11), Rivera-Batiz (1988), discussed the effects of impediments to trade in differentiated goods on the regional distribution of economic activity.

In the policy debates on actual economic integration, such as Europe 1992 or North American Free Trade Area (NAFTA), much has been discussed on the possible impacts on smaller economies. Many argue that, incorporated into a larger market area, small economies can enjoy all the benefits of economies of scale, and become main beneficiaries of economic integration. Others believe, however, that economic integration and free movement of labor and capital leads to a concentration of economic activities into the center, leaving peripheries underdeveloped. The above analysis suggests that both arguments have some theoretical merits and we need much detailed information about the process of economic integration in order to determine the impacts. See Krugman and Venables (1990) and Krugman (1991) for a further exploration on this issue.

4. Complementarity and Agglomeration

The regional disparities caused by migration of workers represent just one example of agglomeration phenomena, more general patterns that we observe everywhere in the real world. For instance, many industries tend to concentrate into a few areas within a country. On a much smaller scale, we see that retail stores and restaurants tend to cluster together in certain sections within a city.
Some sort of complementarity is obviously important for explaining retail store clustering, but we cannot entirely attribute it to the physical characteristics of products sold by these stores. True, we often observe that restaurants and theaters tend to cluster together, as they offer complementary services. In more extreme cases, such as the nuts and bolts or the left and right shoes, complementarities are so strong that they should be sold together in the same store. What is less obvious is that stores that sell very similar products and hence compete directly for customers also cluster together. Examples abound, such as automobiles dealers, bookstores, camera, electronics, and furniture stores, hair dressers, etc. Why do not these stores spread geographically? The need to share the common infrastructure may be a reason in some cases, but the universality suggests that there is something more that make these stores cluster together.

A slight extension of the model in the previous section, taken from Matsuyama (1992c), helps to explain why stores that sell similar products cluster together. The only difference is that the differentiated products, now interpreted as consumer goods, are divided into two groups, 1 and 2; let \( n_i \) denote the product variety offered in group i. The consumer maximizes the following preferences,

\[
V(x_1, x_2) = \left[ x_1^{1-\frac{1}{2}} + x_2^{1-\frac{1}{2}} \right]^{\frac{1}{\epsilon-1}}
\]

where

\[
x_i = \left[ \int_{0}^{n_i} [x_i(z)]^{1-\frac{1}{2}} dz \right]^{\frac{\sigma}{\sigma-1}}, \quad (\sigma > 1)
\]

for \( i = 1 \) and 2. Here, \( V \) aggregates the two composites of differentiated goods, and \( \epsilon \) represents the intergroup elasticity of substitution, while \( \sigma \) may be referred to as the intragroup elasticity of substitution. Note that this model is reduced to the previous one if \( \epsilon = \sigma \). As before, each firm sets the price to be equal to one; and hence all goods in the same group are produced by the same amount, \( x_i(z) = x_i \). This implies that equations (5), (7), and (8) hold for each group. As the relative
demand of the two composites satisfies $X_1/X_2 = [P_1/P_2]^{-\epsilon}$, the ratio of the gross profit in the two groups is equal to

$$\frac{\bar{\pi}_1}{\bar{\pi}_2} = \frac{\bar{\pi}_1}{\bar{\pi}_2} \left[ \frac{n_1}{n_2} \right]^{\frac{\sigma}{\sigma-\epsilon}} = \left[ \frac{P_1}{P_2} \right]^\epsilon \left[ \frac{n_1}{n_2} \right]^{\frac{\sigma}{\sigma-\epsilon}} = \left[ \frac{n_1}{n_2} \right]^{\frac{\sigma}{\sigma-1}} .$$

This shows that the relation between the profit level and product variety depends on the relative magnitude of $\epsilon$ and $\sigma$. For example, suppose that group 1 consists of the restaurants and group 2 retail stores. A pair of restaurants or a pair of stores are much closer substitutes than dining and shopping ($\epsilon < \sigma$). Then, the profit level is negatively related to the variety. If there are too many restaurants and a few stores in the city, restaurants will close down and new stores will open in the long run. Entry and exit processes equalize the numbers of the two types of establishments.

On the other hand, suppose that there are two streets in the city and products are grouped according to their location. It is costly to move back and forth between the two streets, but ex ante the consumers are almost indifferent between the two locations. Then, $\epsilon$ is close to infinity, so that $\epsilon > \sigma$ and hence the profit level is positively related to the number of shops. Entry of a new firm, by attracting more customers, would benefit the existing firms in the same street. This introduces a complementarity in the locational decision, and entry and exit processes lead to all stores clustering into a single location.

It should be noted that a clustering of two products can occur in this model even when the rate of substitution between them, $\sigma$, is high. What makes a pair of products complementary to each other is the presence of a third alternative, rather than the physical characteristics of the two products. This point was made by Hicks and Allen (1934). They criticized the notions of substitutes and complements given by earlier writers and proposed a new definition based on the property of market demand in the presence of many commodities.
5. **Circular and Cumulative Causation in Growth and Development**

The complementarity in the entry processes and associated expanding product variety is also useful for understanding some fundamental problems of economic growth and development. Satisfactory treatments of these issues, of course, require a dynamic model, and in fact, the literature has evolved primarily by extending the dynamic monopolistic competition model by Judd (1985). None the less, I will refrain from developing dynamic models, as the static framework can go a long way to illustrate the main ideas in the literature.

**Underdevelopment Traps**

One critical aspect of development process is that productivity growth is realized through an ever greater indirectness in production. One of the main obstacle to economic development is that complicated technologies often require a variety of nontradeable inputs and producer services. In underdeveloped regions, the lack of local support industries force the use of relatively simple production methods in downstream industries. This in turn implies the small market size for specialized inputs; The lack of local demand prevents the network of support industries from springing up in the region. Thus, the two factors, the lack of demand and the lack of support industries, are mutually interrelated. Not only the division of labor is limited by the extent of the market, but also the extent of the market is limited by the division of labor. Such a circular causation creates underdevelopment traps. Of course, the circularity does not always imply a vicious circle. If regions acquire more than a critical mass of support industries, the very fact that the relation is circular generates a virtuous circle. Over time, the division of labor becomes far more elaborate, the production process more indirect, involving an increasing degree of specialized inputs. Through such a cumulative process, the economy experiences productivity growth and a rising standard of living. In the presence of complementarity, nothing succeeds like success, and poverty becomes its own
cause.

Again, a slight extension of the model of section 3, taken from Matsuyama (1992c, Sec. 3), helps to capture this idea. Let us now suppose that a single consumption good is produced with the following constant returns of scale production function,

\[ C = F(X, N) \]

where \( N \) is labor input. This specialization allows the final goods industry to substitute between labor intensive and intermediate inputs intensive technologies. The relative demand for \( N \) to \( X \) is given by an increasing function of \( P \),

\[ \frac{N}{X} = \Phi(P), \tag{10} \]

with the elasticity of substitution between \( X \) and \( N \) being equal to \( \epsilon(P) = d\log\Phi(P)/d\log P \). The pricing behavior of monopolistically competitive firms is the same as before, so that \( p(z) = 1 \) for all \( z \) and all varieties are produced by the same amount; hence (5) and (7) remain valid. The labor market condition now becomes, instead of (6),

\[ n(ax + F) + N = L. \tag{6'} \]

For any given \( n \), (5), (6'), (7), and (10) can be solved for \( x \). From (8), the gross profit per firm satisfies

\[ \pi(n) = \frac{L - nF}{(\sigma - 1)n + \sigma \Phi \left( \frac{1}{1-\sigma} \right) \frac{\epsilon}{\sigma - 1}}. \tag{9'} \]

This shows that, with a large \( \epsilon(P) \), the profit function can be increasing in \( n \). For example, suppose that \( F(X,N) \) is a CES, with \( \epsilon(P) = \epsilon \). Then, \( \Phi(P) = \beta P^\epsilon \) and the profit function has a single peak when \( \epsilon > \sigma \). In this case, there are three equilibria, as shown in Figure 3. The middle equilibrium, \( S_M \), represents the threshold level below which the firms make losses. If the economy is slightly above
the middle equilibrium, there is an inducement to start up firms. The profit per firm rises with the number of firms around the middle equilibrium, which make it unstable. The other two equilibria, \(S_0\) and \(S_{H_0}\), are both stable. In the lower range, the limited availability of specialized inputs induces the final goods sector to use relatively labor intensive technologies, which implies the small market size for inputs producers. No firm is able to stay in the market, and \(n = 0\), or \(S_0\), represents a state of underdevelopment toward which the economy gravitates. The higher level equilibrium, \(S_{H_0}\), on the other hand, is characterized by a wider range of intermediate inputs and a higher share of the intermediate inputs sector in GNP.

More generally, this model could have an arbitrary number of stable equilibria, as the property of the constant returns to scale imposes very few restrictions on the relative factor demand \(\Phi(P)\). An example can easily be constructed by allowing the final goods sector to have access to a finite number of Leontief (i.e., the fixed coefficient) technologies. Furthermore, it can be shown that the equilibrium per capita consumption is positively related to the equilibrium product variety. The model is thus consistent with the idea of the stages of economic development.

The multiplicity of stable equilibria arises in this model because the benefits of a new input are not completely appropriated by the firm that introduces it. With an increasing variety of inputs, the entry induces the final goods sector to switch to more intermediate inputs intensive technologies, thereby generating demand for other inputs producers. No individual firm, however, does not take into account such pecuniary externalities. Of course, a coordinated, simultaneous entry of firms would solve the demand spillover effects, making it possible for the economy to jump from \(S_0\) to \(S_{H_0}\). However, this is partly due to the static nature of the model. In an explicit dynamic setting, where starting up new firms require reallocation of the current resources from production and the benefits of productivity growth are realized only in the future, Ciccone and Matsuyama (1992) shows that the resource constraint makes the coordinated entry unprofitable, and the economy cannot escape from
the state of underdevelopment; thus \( n = 0 \) becomes a poverty "trap."

It is also worth pointing out that, at a more formal level, the mechanisms generating multiple equilibria in this model are similar to those discussed in the previous section. The possibility of substitution between labor intensive and intermediate inputs intensive technologies makes intermediate inputs complementary to each other, even when the rate of substitution between two intermediate inputs is high.

**International Economics**

The above model also suggests that international trade could be responsible for uneven development. To see this, consider a following model of a small open economy, adopted from Rodríguez (1993). There are two tradeable consumer goods, A and B, and the consumer's preferences are given by \( C_A^\alpha C_B^{1-\alpha} \). Both consumer goods sectors are competitive, and the production function of sector A is \( X \), while that of B is \( N \). In the absence of international trade, this is analytically equivalent to an economy of a single final goods industry with the production function \( F(X,N) = X^n N^{1-n} \). The relative demand is then \( \Phi(p) = (1/\alpha - 1)p \) and the profit function is a decreasing function of \( n \), as depicted by the dotted curve in Figure 4. There is a unique stable equilibrium, in which the economy produces both consumer goods. Some algebra shows that the profit function is

\[
\pi(n) = \frac{\alpha}{\sigma - \alpha} \left( \frac{L}{n} - F \right),
\]

and the equilibrium product variety is

\[
n = \frac{\alpha L}{\sigma F}.
\]

Suppose now that the economy trades in the world market, where the relative price of the two tradeable goods is exogenously given and equal to one. The possibility of trade makes the two
consumer goods perfect substitutes for the competitive firms. Thus, the economy specializes in A when \( P = n^{(1-\sigma)} < 1 \) or \( n > 1 \); it specializes in B when \( n < 1 \). The gross profit is thus given by (6) when \( n > 1 \), but is equal to zero if \( n < 1 \). If \( L > \sigma F \), there are two stable equilibria, \( n = 0 \) and \( n = L/\sigma F \), with \( n = 1 \) being the threshold level, as shown in Figure 4. In this example, international trade creates an underdevelopment trap. Furthermore, if \( aL > \sigma F \), the autarky would help to generate the critical mass of support industries, so that a temporary isolation would help the economy to escape the underdevelopment trap. The model thus captures an element of the views expressed by Baran (1957), Myrdal (1957), and most notably by Prebisch (1950).

*Sustainable Growth*

Note that in the previous model, the cumulative process of growth and development ultimately peters out. This is because the resource constraint eventually becomes strong enough to counteract demand complementarities. In industrialized countries, however, all indices seem point steadily upwards. On the average and in the long run there are no signs of a slacking of the momentum of economic development in these countries. The recent literature on endogenous growth showed that the complementary in the technologies of entry processes itself may be important for sustainable growth.

Again, the model of section 3 can be used to illustrate the basic idea. As shown in Figure 2, the benefit of entry, the gross profit, declines with new entries. In order to offset the tendency of smaller profits, the cost of entry must also decline. One way of doing this, taking from Rivera-Batiz and Romer (1991) and Barro and Sala-i-Martin (1992), is assume that starting up new firms requires \( F \) units of the final good, instead of labor. The idea is that new generations of computers would help scientists to design new products, hence this model is called "the lab equipment model" by Rivera-Batiz and Romer. Under this specification, the entry activity also benefits from increasing
returns due to specialization, which introduces a complementarity in the entry process and makes the
growth sustainable.

As labor is used only in manufacturing of inputs, the labor market condition becomes simply
\( n \alpha = L \). The gross profit is thus \( \pi = x/\sigma = L/(\sigma-1)n \), which is still declining in \( n \). The cost of
entry, on the other hand, becomes \( PF \), so that the entry of new firms would continue as long as

\[
\pi(n) = \frac{L}{(\sigma-1)n} \geq n^{1-\sigma}F, \tag{9*}
\]

or

\[
\frac{L}{(\sigma-1)} \geq n^{2-\sigma} \cdot F.
\]

If \( \sigma = 2 \), then both the benefits and costs of entry fall at the same rate, and the inducement to start
up new firms always exist as long as \( L > (\sigma-1)F \). If \( \sigma < 2 \), the case shown in Figure 5, the cost of
entry falls faster than the benefit. This creates the threshold for development, but the cumulative
process of growth, once started, will never stop.

An alternative way of generating sustainable growth, taken from Romer (1990) and Grossman
and Helpman (1992, Ch.3), is to introduce a learning-by-doing in the process of entry, while
maintaining the assumption of the fixed cost being paid by labor. More specifically, let us suppose
that the amount of labor required in start-up operations declines with the number of firms, say, \( F/n^\lambda \).
This is to say that the stock of knowledge useful for setting up new firms is increasing in \( n \). The idea
is that entrepreneurs starting new businesses or scientists designing new products can learn from the
experiences from the past, hence named "the knowledge-driven model" by Rivera-Batiz and Romer.
The labor market clearing condition is now
\[ L = n \left[ a \times + \frac{F}{n^\lambda} \right]. \]

By substituting this expression in (8), one can show that the entry of new firms would continue as long as

\[ \pi(n) = \frac{1}{\sigma-1} \left[ \frac{L}{n} - \frac{F}{n^\lambda} \right] \geq \frac{F}{n^\lambda} \]

or

\[ \frac{L}{\sigma F} \geq n^{1-\lambda}. \]

If the labor productivity in the start-up operations is proportional to the number of firms \((\lambda = 1)\), as assumed in Romer, and Grossman and Helpman, then there is always an incentive to start up new firms and introduce new products as long as \(L > \sigma F\). If the learning-by-doing or knowledge-spillover effect is stronger \((\lambda > 1)\), then the possibility of underdevelopment traps arises, but, again, the cumulative process of growth, once started, will never stop.

6. **Concluding Remarks**

In recent years, monopolistic competition models have frequently been applied in the context of macroeconomics, international and interregional economics, as well as growth and development. The three features of monopolistic competition, the monopoly power of differentiated goods producers, the lack of strategic interactions, and the explicit analysis of entry and exit processes, make it a useful framework within which to examine the aggregate implications of monopoly distortions, increasing returns and expanding product variety. In this paper, I have presented a highly selective review in this area, with special emphasis on the complementarity and its role of generating multiplier processes, agglomeration, underdevelopment traps, regional disparities, and sustainable growth, or
more generally, what Myrdal (1957) called the "principle of circular and cumulative causation."

I should point out that monopolistic competition is not the only way of modelling complementarity. As the self-adjusting nature of the market force in the standard neoclassical paradigm is due to its efficient resource allocations, any departure from the standard paradigm would increase the chance of generating complementarities, thereby introducing the instability and cumulative processes. For example, it is well-known that more decentralized trading processes, instead of well-organized Walrasian markets, creates complementarities, which have been used to explain underemployment and business cycles (Diamond 1982; Diamond and Fudenberg 1989; Howitt and McAfee 1992; Mortensen 1989, 1991), and the universal adoption of a single medium of exchange (Kiyotaki and Wright 1989; Matsuyama, Kiyotaki, and Matsui, 1992). Recent work on asymmetric information also demonstrated that information externalities generate complementarities and cumulative processes and they are used to explain a variety of phenomena, such as custom, fashion, and fads (Banerjee 1992, Bikhchandani, Hirshleifer, and Welch 1992), market crashes (Caplin and Leahy 1992), and racial segregation (Coate and Loury 1991).

Or for that matter, pure technological externalities, either static Marshallian external economies or dynamic learning-by-doing effects, would be often enough to generate complementarities. Many existing studies (including my own) used external economies to model a variety of phenomena discussed in this paper, while maintaining the other assumptions of the neoclassical paradigm; see, for example, Arthur (1990), Faini (1984), Krugman (1981, 1987), Lucas (1988), Matsuyama (1992d) for international and interregional inequalities and industrial localization; Azariadis and Drazen (1990), Durlauf (1990) and Matsuyama (1991) for underdevelopment traps; Lucas (1988) and Romer (1986) for sustainable growth. Even some monopolistic competition models, such as the knowledge-driven models of sustainable growth discussed in the previous section, rely on pure externalities. It is hard to deny that pure technological externalities resulting from knowledge
spillovers are important in explaining some of these phenomena. But at the same time, some would feel that assuming externalities in the standard paradigm would be a cheap way of generating complementarities. Krugman (1991), for example, argues against relying too much on assuming pure externalities. My own view on this matter is that models with pure externalities are often useful and convenient for exploring the consequences of complementarities, but should be taken at most as a reduced form that is meant to capture some underlying mechanisms generating complementarities. And any result, particularly on the effects of policies, needs to be interpreted with great caution. For any change in the environment might also affect the nature of complementarities itself, as illustrated by a couple of examples in this paper.
References:


Startz, R., "Monopolistic Competition as a Foundation for Keynesian Macroeconomic Models,"


Figure 1

\[ \frac{A_2}{1-\mu_2 \alpha_2} \]

\[ \frac{A_1}{1-\mu_1 \alpha_1} \]

eq. (3)

eq. (4)
Figure 2
Figure 3
Figure 4
Figure 5

\[ \frac{L}{(\sigma - 1) n} \]

\[ \frac{1}{n^{1-\sigma} F} \]

\[ \left[ \frac{(\sigma - 1) F}{L} \right]^{\frac{\sigma - 1}{2-\sigma}} \]