

Increasing Returns: a historical review

*Sylvia Beatriz Guillermo Peon*¹

Introduction

The focus of this paper is to analyze the evolution of the “Increasing Returns” issue within the context of economic theory, and to present a brief analysis of its economic implications. First of all, it is important to remember that a technology exhibits increasing returns to scale, if a proportionate increase in all inputs allows for a more than proportionate increase in outputs; this implies a decreasing average cost curve.

In the wide economic literature we can find a great number of studies involving the increasing returns issue, but it is reasonable to think that it is not possible to present an analysis of all of them in just one paper. Then, I just present here those studies which, in my opinion, represent the most important ones. In the first part of the paper, a historical background is presented. I will go over the work of Adam Smith (1876), Alfred Marshall (1890) and Allyn Young (1928). These economists have been considered as the predecessors of the theory on increasing returns, and an important characteristic of their studies is the fact that they visualize the process of division of labor as the main reason why we observe technologies that

exhibit increasing returns to scale.

Adam Smith placed the division of labor at the forefront of his discussion of economic growth and progress. His explanation about how and why nations are or can become wealthy, is based on the concept of the division of labor, which is considered by him, as the main source of increasing returns. Adam Smith also established a theorem that relates the division of labor to the extent of the market, which can be considered as the explanation of why the division of labor determines a nation’s relative productivity.

Alfred Marshall is an economist from the second half of the 19th century who fully appreciated the importance of the division of labor. He devoted no less than three chapters to the division of labor in his *Principles of Economics* (1890, Book IV, chapters 9-11), not only covering most of the points traditionally dealt with this issue, but often introducing modifications.

¹ I would like to express my gratitude to Axel Leijonhufvud, Professor at UCLA Economics Department, for his valuable comments and corrections on this paper.

For example, Marshall discounted detrimental social consequences from monotonous work. Likewise, he extended Babbage's "principle of economy of skill", to economy of machinery and materials, used it as a major explanatory factor for the localization of specialized industry, and made it the chief advantage of large-scale production. Marshall also made an important distinction between internal and external economies.

Allyn Young, was the first economist who made Adam Smith's theorem regarding the division of labor limited by the extent of the market, the central theme of his work. Young's study was concentrated on two interdependent matters: growth of indirect and roundabout methods of production and the division of labor (or increased specialization) among industries. He visualized the division of labor as a cumulative, self-reinforcing process because every reorganization of production, sometimes narrowly described as a new invention, involves fresh application of scientific progress to industry. Besides this, one of the most important conclusions of Young's work about increasing returns is the fact that he considers operations of individual firms as limited, but certain roundabout methods of production become feasible and economical when their advantages can be spread over the output of the whole industry. The scale upon which the firms in the new industry are able to operate, is the secret to achieve economies of scale.

In the second part of this essay, I will present an analysis of the most important theories, model and applications of increasing returns that have been developed in the last three decades. Then, I introduce the

studies of Kenneth Arrow (1962), Nicholas Kaldor (1972) and Paul Romer (1987). Arrow and Romer developed a formal model of increasing returns processes. In Arrow's work, learning-by-doing is one of the reasons giving rise to dynamic economies of scale. Specifically Arrow suggests an endogenous theory of the changes in knowledge, which underlies intertemporal and international shifts in production functions.

Nicholas Kaldor was one of the few major economists who took up Young's challenge. Kaldor's work on this issue relates to international trade. He made a severe critique to the assumptions on which international trade rested, saying that, there are some stylized facts that show the existence of increasing returns technologies in some countries engaged in international trade, and these facts do not coincide with the traditional assumptions made by the international trade theory to explain the advantages and benefits of trade. Therefore, he explains that not all the countries engaged in international trade necessarily benefit from this trade. Specifically Kaldor mentions that, when two countries are engaged in trade, and one of them faces an increasing return to scale production function, and the other faces a nonincreasing returns production function, the latter could end up being a much poorer country than without trade. One of the most important arguments that Kaldor suggested was that, faster growth is derived from faster growth in the manufacturing sector, partly because of the cumulative features which link the growth of manufacturing to growth of labor productivity via economies of scale (or the notion of increasing returns developed by Young from the division of labor).

Paul Romer's work suggests a model of increasing returns arising from specialization in production. His main conclusion is that, when increasing returns are present in an economy as consequence of high degree of specialization in production processes, the social planner solution to the problem of intertemporal optimization of welfare, will produce to have a higher rate of investment and a higher rate of growth than the decentralized, competitive. So all individuals can be made better off by an agreement.

Part I Historical Background

Adam Smith: The division of labor as an important source of *increasing returns*.

Adam Smith in the *Wealth of Nations*, made the division of labor his grand theme. For him, division of labor is the great cause of its increased powers, and therefore an important cause of economic progress. On the other hand, the role of machinery as a source of increasing returns is also treated by Smith but as secondary and subsidiary to increasing division of labor. Smith's division of labor, in his theory of production, has been seen by many authors as the starting point of the concept of what is called "technological change coefficient" under the modern production theory, or as economies of scale property of the production function.

But, what could we say about the definition of the division of labor? The division of labor may be defined as the division of a process or employment into parts, each of which is carried out by a separate person, and in this way, division of labor is associated with labor productivity.

In order to understand how the division of labor makes labor force more productive, Adam Smith explained his famous pin-making example. In this illustration, two modes of organizing production were contrasted: craft production and factory production (as named by Axel Leijonhufvud). In crafts production, each craftsman sequentially performs all the operations necessary to make a pin. In factory production each worker specializes in one of these operations:

One man draws out the wire, another straightens it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on, is a peculiar business, to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations.

If the pin-making is carried on the craft production way, each craftsman producing a pin requires to do all the same sequence of operations by its own, and at his own pace, since the individuals differ in skill across operations. So, by simply rearranging the work in some given workshop, output is produced by a team in which each individual performs one task and this work has to be done at the pace of the team. This reorganization in the way of producing is translated in a significant increase in output. According to Smith, the advantages of the division of labor are due basically to three circumstances: first to the increase in dexterity in every particular workman; second-

ly to the saving of the time which is commonly lost in passing from one species of work to another; and lastly to the invention of a great number of machines which facilitate labor, and enable one man to do the work of many.

The factory production system also makes supervision of work effort easier, and even without changing the engineering prescriptions of the operations performed, even more, without changing the tools used and the people involved we should expect a large increase in productivity from this reorganization of work.

The economies achieved by switching from crafts to factory production arise from increased division of labor. In Smith's example the conversion takes us from individual production to team production. But, there are some aspects that we should take into account. While the specialization of labor in team production requires standardization of product, under craft production, in contrast, the different skills of individual artisans, will be reflected in nonstandard output. Also, serial production requires coordination of activities, and maybe one of the most important features, is that labor of individual workers become complementary inputs, in the sense that, if one work station on an assembly line is unmanned, total product goes to zero.²

In general, when we talk about the production of a good, the conversion from crafts to factory production will present opportunities to economize on inputs. The switch is capital-saving (i.e. instead of re-

quiring n hammers for say, N artisans, we just require hammers for those stages of the process in which the hammer is used in that special task). And also, this switch from craft production to factory production, will save on human-capital, since no worker needs to possess all the skills required to make a pin from the beginning to the end.

What gives rise to the division of labor in the firm? Smith says that, division of labor, from which many advantages are derived, is not originally the effect of any human wisdom, with the purpose of opulence:

It is the necessary, though very slow and gradual consequence of a certain propensity in human nature which has in view no such extensive utility; the propensity to truck, barter, and exchange one thing for another.

So, it is the **power of exchanging** (or trade) that gives rise to the division of labor, and then, as Smith says, the extent of this division must always be limited by the extent of that power, or in other words, by the extent of the market. When the market is very small, no person can have any encouragement to dedicate himself entirely to one employment in which he cannot get anything in exchange to all the product that exceeds his consumption. From this part we can see that Smith puts a very special attention to the power of trade. When he says that the power of exchange gives rise to division of labor and hence to increasing returns, he is implicitly acknowledging the existence of mutual gains from trade.

The division of labor depends on the extent of the market, and so do the scale economies that can be realized. These econ-

² Cf. Axel Leijonhufvud, 1986; *Capitalism and the Factory System*. In R. Langlois (ed.), *Economics as a Process*. Cambridge University Press.

omies of scale, as seen by Smith, are the consequence of an increased **vertical** division of labor. As he thought, when the extent of the market grows, opportunities arise for further subdivision of the production process into a greater number of serial tasks; therefore vertical division of labor results from an increasing returns to scale technology.

Although the proposition that the division of labor is limited by the extent of the market³ was not strictly demonstrated by Smith, this idea was utilized by many economists like Young (1928), Coase (1937) and Stigler (1951), to provide the fundamental features of a theory of vertical integration and production roundaboutness. On the other hand, this proposition regarding the limits of the division of labor, has some important implications which have been also mentioned by Smith in his "Wealth of Nations". The first implication is related with the profits rate in the industry. Smith says that small economies devote most of their resources to the agriculture, while large economies specialize in industry, because the latter affords a greater degree of division of labor. For exactly the same reason, increases in market size decrease the price of industrial products relative to the primary products, and as a consequence the profit rate in industry declines.⁴ The second implication is related to the benefits of trade. In Smith's view, trade increases the market size and allows each trader to specialize and reap the benefits of increased division of labor. Trade is therefore beneficial to all parties involved since it increases

real income of all classes, and therefore should not be restricted by governments.⁵

At this point, it is important to ask ourselves, how could we make precise the relationship between extended specialization and economic productivity? That is, why and how does extended division of labor translate into a more highly valued bundle of goods and services? We can say that it does so because the economy, treated as a whole, exhibits generalized increasing returns which become merely a technical way of stating Smith's central proposition. The output-input ratio increases as the size of the economic network (measured by the quantity of inputs) increases because of the extension in input specialization. Hence any defined increase in quantities of inputs will generate a disproportionately larger increase in quantities of outputs. Returns to the overall scale of the economy are increasing.

Axel Leijonhufvud (1986) makes in my opinion, a clear analysis of the social consequences in the 19th century factory system of Smithian vertical division of labor that arise from the competitive impetus to exploit the economies afforded by this division of labor:

1) When labor is subdivided vertically, less skill is required, less versatility as producer is acquired by the individual worker. The use of child labor at some work stations often becomes feasible.

2) No normal prospect of promotion or improvement in social status is to be expected; the unskilled workman does not become a master of his guild by sticking to his job for many years.

³ The Wealth of Nations, Book I, Ch. III.

⁴ The Wealth of Nations, Book I, Ch. XI, and Book III, Ch. I.

⁵ The Wealth of Nations, Book IV, Ch. II and Book I, Ch. II.

3) More discipline is required; workers cannot work at their own pace, they have to be on time; random absenteeism must be subject to relative severe sanctions.

4) “Alienation from the product”: No worker can take personal pride in the output or its quality.

In addition to these four points I will include among the consequences (in this case not social consequences) of the division of labor one more that is considered the next step in its evolution: the mechanization, which is also analyzed by Leijonhufvud (1988). The continued subdivision of labor results in operations so mechanical that a machine can do them, and do them both faster and better. The exploitation of the economies that open up as the extent of the market grows, produces increased functional differentiation of both capital equipment and labor. But the implications for capital and labor are not symmetrical. Inputs tend to be complementary to one another (i.e. the assembly line stops if one worker is missing or one machine breaks down). The typical machine is highly specialized, that is, dedicated to particular tasks in the manufacture of a particular product. It may have no alternative uses but is, on the other hand, not quickly or easily replaced: it has a thin market. The typical factory worker works at specialized task, but an unskilled one has lots of alternative jobs for which he could easily and quickly qualify, and hence the unskilled worker becomes an easily replaced factor: his market is a thick one. Because of the returns to scale, the enterprise typically earns a monopoly rent, and because inputs are complementary this is a joint rent.

According to Leijonhufvud, the joint rent creates a distributional problem which

must be solved in order to exploit the returns to scale for the production factors. Complementarities among inputs mean that marginal productivities are undefined and give no guidance to a fair distribution. Division of the joint rent becomes a bargaining problem that is only partly determined by the cooperating inputs alternative opportunities in outside markets. To stabilize the cooperative arrangement required for the exploitation of the economies of scale, the best that can be done, very often leaves us with the management facing unionized labor. The firm, therefore is created to control all machines that are complements, and have thin outside markets.

Next, I will analyze the ideas of Marshall regarding increasing returns, and through this analysis we could realize that Smith’s key point of division of labor, also prevails in Marshall’s work.

Alfred Marshall: The specialization and mechanization of work, and the increase in the scale of *production*.

Alfred Marshall, in his famous book *Principles of Economics*, seems to agree with Adam Smith, with respect to the fact that division of labor and specialization in lower grades of work (i.e. manual tasks), increase efficiency in production. Marshall says that the extreme specialization also reduces the action of the workers to routine, and it is in this stage, that the work can be taken over by machinery. The improvement in machinery and the growing division of labor, have gone together, and are in some measure connected. As Smith pointed out, Marshall says that

it is the largeness of the market, the

increased demand for great number of things of the same kind, that leads to subdivision of labor.

Besides, Marshall makes a brief explanation of how mechanization arises from the division of labor, saying that the power of machinery to do work that requires too much accuracy to be done by hand, displaces the purely manual skill, this is one of the consequences of the division of labor. On the other hand, the introduction of machinery as substitute to pure manual and monotonous work, leads to some advantages that, according to Marshall, one of the most important, is the reduction of the 'evil of monotonous work':

For those trades in which the work is most subdivided are those in which the chief muscular strain is most certain to be taken off by machinery; and thus the chief evil of monotonous work is much diminished. As Rochester says, it is monotony of life much more than monotony of work that is to be dreaded: monotony of work is an evil of the first order only when it involves monotony of life.

For Marshall, the use of machinery creates a tendency to increase the scale of manufactures and to make them more complex, and therefore to increase the opportunities for division of labor of all kinds. Marshall also made an important division of the economies arising from an increase in the scale of production, into two classes:

a) Those dependent on the general development of the industry, called external economies, and

b) Those dependent on the resources of the individual firm, on their organization and the efficiency of their management, called internal economies.

In Marshall's idea, the chief advantages of production on a large scale are: economies of skill, economies of machinery, and economies of materials. For example, he says that, when a hundred sets of furniture, or of clothing, have to be cut out on exactly the same pattern, it is worthwhile to spend great care on so planning the cutting out of the boards or the clothes, that only a few small pieces are wasted. This is properly an economy of skill since one planning is made to suffice for many tasks.

With respect to economies of machinery, it is well known that small manufacturers often cannot afford to buy machinery, and usually a small manufacturer does not have enough space for the specialized machinery. By contrast, large establishment is able to afford highly specialized machinery, and this also applies with highly specialized skill. On the other hand, Marshall says that a large business buys in great quantities and therefore cheaply; it pays low prices and saves on transportation in many ways. The large manufacturer has a much better chance than a small one has, to select men with better abilities for work. The increase in the size of firms (increase in the scale of production) requires increasing skills on the part of the people who are managing the business, which also leads to a highly developed industrial organization, and in turns adds much to the collective efficiency of capital and labor. The increase in the scale of business increases rapidly the advantages of the firm over its competitors, and lowers the price at which this firm can sell.

What is important to Marshall, is that the aggregate scale of production, results in growth of wealth, but also, an increase of this types of economies has a very important influence in the way of determining the supply price of a commodity. The general argument made by Marshall in his Principles of Economics about this issue, shows that an increase in the aggregate volume of production of anything, will generally increase the size, and therefore the internal economies possessed by such firm; that it will always increase the external economies to which the firm has access, and thus will enable it to manufacture at a less proportionate cost of labor and sacrifice than before:

We say broadly that the part which nature plays in production shows a tendency to diminishing return, the part which man plays, shows a tendency to increasing return. The law of increasing return may be worded thus: An increase of labor and capital leads generally to improved organization, which increases the efficiency of the work of labor and capital.

In his same book, Marshall defines:

Increasing return is a relation between a quantity of effort and sacrifice on the one hand, and a quantity of product on the other.

We should understand this sentence, in the sense that there is a trend toward a smaller quantity of effort and sacrifice per unit of product. Marshall also analyses the relation of industrial expansion and social

welfare. He mentions that an increase on industrial efficiency generates an increase of wealth, in the sense that people face an increase of the means of satisfying human wants.

For Marshall (as well as for Smith), increasing returns are the result of the division of labor and specialization, which can be translated into efficiency: decreasing costs. But, it is important to mention that the famous statement of Alfred Marshall regarding increasing returns, is the fact that increasing returns tends to monopoly, because

...some producers get ahead of their rivals and gain a cumulative advantage over the others whom they will drive out of business.

Hence, for Marshall, increasing returns (or falling marginal costs) could not exist under conditions that prevail in a competitive market. Moreover, Marshall recognized that the industries in which particular processes exhibit increasing returns to scale must rapidly become monopolized. Although Marshall did not explicitly acknowledge the failure of competitive processes to generate the allocation that will produce maximal value, it is Kenneth Arrow and also it is Paul Romer who realized this fact as result of their dynamic models. I will go through this point in the second part of the paper.

Allyn Young: The roundabout methods of production *and the specialization among industries*

In his 1928 paper "Increasing Returns and

Economic Progress”, Allyn Young revived Adam Smith’s central proposition on the **division of labor**, and he sketched out the relationship between economic progress and the presence of economywide increasing returns. Young focused his paper on two related aspects: the role of indirect or roundabout methods of production and the division of labor among industries. According to Young, with the division of labor a group of complex processes is transformed into a succession of simpler processes, some of which are possible with the use of machinery. In the use of machinery and the adoption of indirect processes there is a further division of labor, the economies of which are again limited by the extent of the market.

For Young, Mr. Ford’s method of producing automobiles would be absurdly uneconomical if its output were very small, and would be unprofitable even if its output were what many other manufacturers of automobiles would call large. According to Nicholas Kaldor (who was an important follower of Young’s ideas) we can interpret this to mean that the extent to which capital is used in relation to labor is predominantly a matter of scale of operations. That is, the capital/labor ratio in production is a function of the extent of the market. Then continuing with Young’s ideas, the convenience to keep equipping factories with special appliances for making hammers or constructing specialized machinery for use in making different parts of automobiles, depends again upon how many nails are to be driven and how many automobiles can be sold.

Young also mentioned that the principal economies which manifest themselves in increasing returns, are the economies of

capitalistic or roundabout methods of production (also called indirect methods of production: those processes which are divided into a succession of simpler tasks and include the use of machinery on them). These economies are identical with the economies of the division of labor in its most important modern forms, and also depend upon the extent of the market.

Young says that it is not only the economy of indirect methods of production that faces increasing returns (which is the most obvious to perceive), but also the economies of large-scale operations and mass-production:

...no one can doubt that there are genuine economies to be achieved in the way of simplification and standardization.

For example, he mentioned that in certain industries, productive methods are (in his time) economical and profitable in America, which would not be profitable elsewhere. Taking a country’s economic endowment as given, Young says that the most important single factor in determining the effectiveness of its industry appears to be the size of the market. But, what constitutes a large market? For Young, it is not the area or population alone, but buying power, that is, the capacity to absorb a large annual output of goods. Besides, he explained that the capacity to buy depends upon the capacity to produce, so the size of the market is determined and defined by the aggregate of productive activities, tied together with trade. So, the mutual advantages from trade are also taken into account by Young as it was in Smith’s work. On the other hand, Young visualized the division of labor as a cumulative self-reinforcing process, because:

Every important advance in the reorganization of production, regardless of whether it is based upon anything which, in a narrow or technical sense, would be called a new 'invention', or involves a fresh applications of the fruits of scientific progress to industry alters the conditions of industrial activity and initiates responses elsewhere in the industrial structure, which in turns have a further unsettling effect. Thus, change becomes progressive and propagates itself in a cumulative way.

In Young's ideas, the apparatus which economists have built up for the analysis of supply and demand in their relations to prices does not seem to be particularly helpful for the purposes of analyzing these broader aspects of increasing returns. This could be a damaging conclusion for competitive price theory. Once again, the coexistence of increasing returns economies and a competitive market environment is questioned.

Young also pointed out that industrial operations must be seen as an interrelated whole, in the sense that the supply of a consumption good (produced by some industry) is related to the demand of intermediate goods (produced by other industry). He explained this assuming that the economy starts operations of reciprocal demand, when the commodities exchanged are produced competitively under conditions of increasing returns and when the demand for each commodity is elastic, in the special sense that a small increase in its supply will be attended by an increase in the amounts of other commodities which can be had in exchange for it. Under such conditions,

Young explains that an increase in the supply of one commodity is reflected in an increase in the demand for other commodities, and it must be assumed that every increase in the demand will evoke an increase in supply. Since the elasticities of the demand and supply will differ for different products, some industries will grow faster than others. So, even with constant population and in the absence of new discoveries, there are no limits to the process of expansion except the limits beyond which demand is not elastic and returns do not increase.

On the other side, Young tell us that we should take into account the existence of various other factors which reinforce the influences that make for increasing returns. The discovery of new natural resources and of new uses for them, and the growth of scientific knowledge are probably the most important of these factors. Besides, Young suggested that the so called potential market, has acquired a new importance in the planning and management of large industries...

The difference between the cost per unit of output in an industry or in an individual plant properly adapted to a given volume of output and in an industry or plant equally well adapted to an output five times as large, is often much greater than one could infer from looking merely at the economies which may accrue as an existing establishment gradually extends the scale of its operations. Potential demand, then, in the planning of industrial undertakings, has to be balanced against potential economies... The search for markets is partly a matter of augmenting profits by reducing costs.

Increasing returns are often related to the growth of industries, and Allyn Young has not tried to avoid that way of thinking, although he explained that it might be a misleading way of understanding industry growth. Young also says that, with the extension of the division of labor among industries, the representative firm loses identity. Its internal economies dissolve into internal and external economies of the more highly specialized processes, and are supplemented by new economies. Therefore, the division of labor among industries is a vehicle of increasing returns, and it has some advantages of its own which are independent of changes in productive techniques. For example, it allows a higher degree of specialization in management, and the advantages of such specialization are real. Again, this specialization leads to a better geographical distribution of industrial operations, and this advantage is very important. Young mentions that nearness to the source of supply of a particular raw material or to cheap power, or to cheap transport, and nearness to a larger center of population are advantages of specialized industries that should be taken into account. But, according to Young:

the largest advantage of the division of labor among industries, is the fuller realizing of the economies of capitalistic or roundabout methods of production. This should be sufficiently obvious if we assume that in most industries there are effective, though elastic, limits to the economical size of the individual firm. The output of the individual firm is generally a relative small proportion of the aggregate output of an industry.

Then, for this economist, the degree in which the individual firm can develop economies of scale by making its own operations more roundabout, is limited. But, certain roundabout methods become feasible and economical when their advantages can be spread over the output of the whole industry. These potential economies then, are segregated and achieved by the operations of specialized undertakings which taken together constitute a new industry. So, we can see that the scale upon which the firms in the new industry are able to operate is the secret of their ability to achieve economies of scale for the industry as a whole, while presumably these firms are making profits for themselves. Therefore, it is an important conclusion that the scale of the operations of the firms in the new industry is what determines the size of the market for the final products of the industry or industries. And the principal advantage of large-scale operations at this stage is that it again makes methods economical, which would be uneconomical if their benefits could not be diffused over a large final product.

As I mentioned before, one of the major followers of Young's ideas was Nicholas Kaldor. As the next step in the present survey, I will go through some of the most important modern theories regarding increasing returns, among which we can find Kaldor's studies.

Part II Modern theories and models

Kenneth J. Arrow: Learning by doing as a source of increasing *returns*.

In his famous 1962 paper "The Economic

Implications of Learning by Doing”, Kenneth Arrow mentioned that “the experience” is an important factor which determines technological changes, and a very important source of increasing returns in the production function:

...technical change in general can be ascribed to experience; it is the very activity of production which gives rise to problems for which favorable responses are selected over time.

Then, for Arrow, “experience” is a legitimate factor of production, since the productivity of labor inputs tends to grow markedly after the workers become familiar with the production process in which they are engaged.

Motivated by the idea that increases in per capita income cannot be explained simply by increases in the capital-labor ratio, Arrow made an important contribution to the theory of growth: he added the obvious fact that the knowledge is growing in time, and this is an important issue that do not directly contradict the neoclassical view of the production function. In Arrow’s study, the view of economic growth, depends so heavily on an exogenous variable: the quantity of knowledge, which is very difficult to measure.

Arrow suggests an endogenous theory of the changes in knowledge that underlie intertemporal and international shifts in production functions. For Arrow, the acquisition of knowledge is what is usually termed “learning”, and learning is the product of experience, that is learning can only take place during activity. So, the knowledge has to be acquire, and since educational experi-

ences for the same period of time, are different among countries, then the countries will have different production functions, even if they have the same natural resource endowment.

Arrow’s theorems about the economic world presented in his article, differs from those in most standard economic theories, in the sense that profits are the result of technical change, and the rate of investment will be less than the optimum (in a free enterprise system), and also the net investment and the stock of capital become subordinate concepts, with gross investment taking a leading role. The first issue in Arrow’s model is how he measures the “experience” in order to capture the technical advance. The index of experience defined in his paper is the cumulative gross investment (or cumulative production of capital goods), since:

...each new machine produced and put into use, is capable of changing the environment in which production takes place, so that learning is taking place with continually new stimuli.

This at least makes plausible the possibility of continued learning in the sense here, of a steady rate of growth in productivity. Following the models of Solow and Johansen,⁶ in which technical change is completely embodied in new capital goods, Arrow says that under his model’s assumptions, the production process associated with

⁶ Arrow followed the ideas prented by Solow in his 1959 paper “Investment and Technical Progress”, and by Johansen, L. in his 1959 paper “Substitution VS Fixed Production Coefficients in the Theory of Economic Growth: A synthesis”.

any given new capital good is characterized by fixed coefficients, so that a fixed amount of labor is used and a fixed amount of output obtained. In Arrow's model, the output of the economy is a function of the cumulative gross investment, and of the labor force employed:

$$Y = F(G, L)$$

where:
 Y = total output
 G = cumulative gross investment
 L = labor force employed

The functional form for the production functions proposed by Arrow, shows increasing returns to scale in the variables G and L . These production functions are as follows:⁷

$$Y = aG \left[1 - \left(1 - \frac{L}{cG^{1-n}} \right)^{1/(1-n)} \right] \dots\dots\dots(1)$$

for $n \neq 1$, and

$$Y = aG (1 - e^{-L/b}) \dots\dots\dots(2)$$

for $n = 1$, and where

$a > 0$ is a constant which represents the output capacity given by the cumulative gross investment
 $n > 0$

⁷ It should be mentioned here that the functional form proposed by Arrow is related to an equation found in an study of learning curves for airframes. That equation was developed by aeronautical engineers, particularly by T.P. Wright, and it means that the number of labor-hours expended in the production of an airframe, is a decreasing function of the total number of airframes of the same type previously produced ($N_{t+1} = N_t^{-1/3}$).

$$c = b / (1-n)$$

b = some positive constant

From equation (2) we can see that an increase in G with L constant, increases Y in some proportion; a simultaneous increase in L will further increase Y . From equation (1) and if $n < 1$, a proportional increase in L and G , increases L/G^{1-n} , and therefore increases the expression in brackets which multiplies G . A similar argument holds if $n > 1$.

But what consequences arise from this particular production function facing increasing return to scale? The knowledge externality presented in Arrow's model seems to have problems with the neoclassical assumptions about income distribution, because although the idea that growth can occur in the absence of exogenous technical progress is very promising, increasing returns in the production function lead us to a another way of maximizing profits, different from just taking wages and rental rate as given, as would be the case under perfect competition.

If the production function does not satisfy decreasing returns to a factor, then the marginal product curves would either be flat (the case of constant returns to scale), or slope up (the case of increasing returns to scale). In the case of increasing marginal productivity of capital, the point where the marginal product of capital equals the rental rate is a point of minimum profits rather than a point of maximum profits, and a firm in this position would maximize its profits by hiring all of the capital in the economy; that is, the presence of increasing returns will lead to a particular firm becoming a monopoly.

However as Arrow pointed out, the increasing returns do not lead to any difficulty

with distribution theory, since both capital and labor are paid their marginal products. According to Arrow, the explanation is that the private marginal productivity of capital (new investment) is less than the social marginal productivity, since the learning effect is not compensated in the market. In other words, even when the production function of the whole economy faces increasing returns due to the knowledge externality, the individual firms cannot make exclusive the accumulation of knowledge of the whole society, in the sense that if one firm produces an idea, so another firm can copy that idea. It is in this way that private firms face diminishing marginal productivity of capital (and labor), and therefore, each private firm will take factor prices as given, paying to each factor its private marginal productivity.

In 1988, King, Plosser and Rebelo, re-take Arrow's theory of learning by doing, and make an important survey regarding this topic.⁸ These authors present again the idea that the theory of learning by doing reconciles the assumption of increasing returns technology with a theory of distribution based on perfect competition, by drawing a distinction between the production function that faces the society as a whole: the *social technology*, with the production function faced by each individual firm: the *private technology*.

Following the ideas presented by Roger Farmer (1995) in his studies about theory of endogenous growth, we can say that in the theory of learning by doing, labor becomes more productive through time, not because exogenous improvements in technology, but because of the accumulation of knowledge.

⁸ See Farmer (1995).

As society grows, the individuals learn new techniques and their knowledge becomes embodied in human capital, and this human capital is a social process that has effects going beyond the individual's own productivity. In other words there is a knowledge externality, and as society grows, the accumulation of knowledge spreads in a way that cannot be appropriated by private individuals. This is basically the reason why the private technology faces diminishing marginal productivity in factors, and the social technology faces increasing returns.

Arrow's model of learning by doing has also another important economic implication: under his assumptions the presence of learning means that an act of investment benefits future investors, but this benefit it is not paid for by the market. Hence it is to be expected that the aggregate amount of investment under the competitive model, will fall short of the socially optimum level. And it also follows that the gross savings ratio is smaller along the competitive path than along the optimal path. The model of learning by doing represents the starting point of a big change in the way of viewing the causes of growth for an economy, since all growth theories of the 1950s treated technology as exogenous and hence beyond the influence of choices that might be made within the economy. The idea that growth it is not necessarily explained exogenously by technical progress but by endogenous factors such as specialization, was followed by Romer in 1986 and 1987.

Nicholas Kaldor : increasing returns in international trade.

The main work of Nicholas Kaldor regard-

ing increasing returns was related to international trade, and it represents a severe critique to the assumptions in which traditional theory of international trade rests. Both, classical and neoclassical theory, assert that free trade in goods between different regions is always to the advantage of each trading country, and is therefore, the best arrangement from the point of view of the welfare of the trading world as a whole, as well as of each part of the world taken separately. However, for Kaldor this is true only under certain conditions for the trading countries. Those conditions or assumptions are not necessarily true. So for example, when one the trading countries has increasing returns production functions, then the trading advantage no longer exists for both countries:

these propositions are only true under specific abstract assumptions which do not correspond to reality. Under more realistic assumptions, unrestricted trade is likely to lead to a loss of welfare to particular regions or countries, and even to the world as a whole -that is to say that the world will be worse off under free trade than it could be under some system of regulated trade.

As soon as we allow for either diminishing returns and/or increasing returns due to economies of scale, the proposition will no longer hold.

For Kaldor, the traditional trade theory (i.e. Hecksher-Ohlin,⁹ Samuelson¹⁰) rests

⁹ These two Swedish economists, Eli Hecksher and Professor Ohlin, show that under certain as-

on artificial assumptions, which are not always stated or even understood:

- Production functions for different goods are the same in different countries;
- Perfect competition prevails;
- Constant return to scale (homogeneous and linear production functions) for all processes of production.

These two latter assumptions are, in Kaldor's idea, the critical point of this theory in the sense that, if two countries are engaged into free trade, and one of them has a production function facing increasing returns (or falling costs), the price of that good can fall so low that the real income of the other country could shrink, since the addition of its output from exports may not compensate for the loss of output due to import competition. So the country with nonincreasing returns production function,

assumptions, differences in comparative costs between countries can only exist if resource endowment, or factor proportions of the different regions, are different; such differences must be reflected in differences in relative factor prices, and the effect of trade must be to bring relative factor prices into a closer relationship to one another.

¹⁰ Samuelson, in his well known "factor price equalization theorem" carried Hecksher-Ohlin's doctrine a stage further showing that, under certain assumptions, the effect of free trade must be to equalize factor prices in the different participating areas. The implications of these theorems are the following: 1) the free movement of goods is a substitute for the movement of factors; 2) trade implies a tendency to equalization of factor prices, as the free mobility of factors would cause; 3) trade must necessarily reduce the differences in real earnings per capita between the different trading areas, and in favorable circumstances (i.e. identical and well behaved production functions), eliminate them altogether.

could end up by being a much poorer country than without trade (since there would be less employment and less output).

Following Kaldor, businessmen engaged into international trade could never ignore the existence of diminishing costs (increasing returns). It is on account of the economies of large-scale production, that a rising market share means success, and falling market share means trouble. From Kaldor's point of view, this is can be the explanation to what happened as a result of the Industrial Revolution in the 19th century. Areas which were previously isolated, became drawn into the world economy, and as Kaldor says, this does not mean that the enlargement of the markets benefited all participating areas in the same way. The manufacturing industry of Britain received an enormous stimulus through the opening of markets in Europe, America, India and China. But at the same time the arrival of cheap factory-made goods, eliminated local producers who became uncompetitive, and it made these countries "specialize" in the production of raw materials, which however could only offer employment to limited number of workers. As a result of this, the countries dependent on the exports of primary products remained comparatively poor. So the poverty was a consequence, not of low productivity of labor in their export sectors, but of the limited employment capacity of their profitable industries (non increasing returns economies).

Kaldor continues explaining that the polarization process¹¹ was concentrated by

¹¹ This is the way Kaldor refers to the concentration of manufacturing production in certain areas a result of free trade in the field of manufactured goods.

the successful spread of industrialization to other countries. The new techniques developed by the factory system in England, would be sooner or later copied by other countries. And so, they were. In the second half of the 19th century, France, Germany, Italy and many other small countries, began to industrialize behind the protection of a newly established tariff system. It is in this way that the spread of industrialization over wider and wider regions, was concentrated by the "polarization effect", which is nothing other than the inhibiting effect of superior competitive power of the industrially more efficient and dynamic countries, as compared to the others. Hence, for this author, without instruments like tariffs and subsidies, industrialization could never have started in small countries that had trade with other countries facing increasing returns to scale.

Paul Romer: Increasing returns due to specialization.

Romer's 1987 paper describes an attempt to model increasing returns that arise because of specialization. His model ignores increasing returns from investments in knowledge. It focuses exclusively on the role of specialization. The basic idea of his paper is that, if we include the number of intermediate inputs used in the production as an argument of the production function of a final good, the resulting technology is an increasing function of the number of intermediate inputs (that is, the output of the final good increases with the number of intermediate inputs used, *ceteris paribus*). This representation captures the idea that an increase in the degree of specialization in-

creases output. So, the production function specified by Romer's model is like the following:

$$Y(L, \{M, N\}) = LM g(N/LM)$$

where M measures the range or number of intermediate inputs used, and N measures the total quantity of such inputs. If $g(\cdot)$ is a power function, then:

$$Y(L, \{M, N\}) = M^{1-\alpha} (L^{-\alpha} N^\alpha)$$

In this equation presented by Romer, Y appears to exhibit increasing returns to scale even when M and N are not relevant inputs. As a function of labor (L) and the list of intermediate inputs $X_i = N/M$, Y is a concave production function, homogeneous of degree one. Romer's model also tries to capture the idea that fixed costs limit the degree of specialization. A decentralized equilibrium for Romer's economy consists of a continuum of firms in the intermediate goods sector, and an indeterminate number of firms producing final output goods with constant returns to scale production function. The final goods firms are assumed to be price takers in all of their markets. Each of the intermediate input producing firms is the single producer of a particular intermediate input and has power in the market for its specialized good. It is still a price taker in the market for primary input.

The kind of equilibrium that Romer obtains is a monopolistically competitive equilibrium. Some of his assumptions are the following: potential and actual producers of intermediate goods maximize profits taking the downward sloping demand curves for these inputs, and the price of the primary

input (or primary resource) as given. In equilibrium, some intermediate goods are produced, and others are not. All firms in the intermediate goods industry earn zero profits, and the price of these goods is a function of the price of the primary resource. The price of the primary resource is determined by the requirement that profits for the intermediate goods producers must be zero. For a given stock of primary resources, the key quantities to be determined are M (the number of intermediate inputs to be produced) and X_i (the amount of each input).

A curious feature of Romer's model is that, if the function $g(\cdot)$ is a power function, then the quantities from the social optimum problem, coincide with those in the decentralized equilibrium. This result relies crucially on the fact that the stock of primary resource is given. The problem arises when we allow an alternative use for the primary resource, because then, following Romer, the decentralized equilibrium will differ from the first-best social optimum. In particular any model that explains growth by allowing individuals to forego current consumption and accumulate additional units of the primary resource, will necessarily have an equilibrium with less accumulation of this resource, than would be socially optimal.

Then, for the dynamic version of this model, Romer solved for the representative agent's problem allowing the individual to accumulate the primary resource, and assuming a utility function with isoelastic form, and a Cobb-Douglas production function for the consumption good. At time t , the consumer will receive labor income given by the marginal product of labor, times the number of labor units worked, and a rental

income given by the interest rate times the number of capital units rented to the market (where the interest rate is taken as given). So, the consumer chooses how much to consume and the rate of accumulation of the primary resource.

Just as in the Romer's static model, the equilibrium condition in the market with monopolistic competition is that the range of inputs produced at time t , must satisfy $M(t) =$ aggregate savings. Each individual consumer takes the path for $M(t)$ as given, because it depends on the aggregate savings decisions for all consumers in the economy. In this sense, $M(t)$ behaves just like a positive externality.

Romer verified that this equilibrium is suboptimal, since relative to the maximization problem faced by each consumer, a social planner would not take the path of wages or $M(t)$ as given; instead, the planner would take account of the fact that a higher rate of savings leads not only to higher investment income, but also higher labor income. The planner would also produce more output for a given stock of primary resources by setting X^* and M at optimal levels rather than at equilibrium levels. But, according to Romer, these effects cause for social planner solution to have a higher rate of investment and a higher rate of growth. All individuals can be better off (by an agreement).

From Romer's point of view, the analysis of this equilibrium, resembles one with a positive externality, and this apparent "external economy" associated with specialization, is closely related to the intuition behind Marshall's term. Although the model is not one with a true positive externality, it behaves exactly as if one were present. The

only intervention needed to achieve the optimum in this special case is a subsidy of savings. Therefore, one of the main conclusions from Romer's work is the fact that, under the presence of increasing returns (in this case as a result of specialization), the market resource allocation is not an efficient allocation, and hence an intervention is needed to achieve an optimum.

Summary and Conclusions

This paper has included some of the central contributions to generalized increasing returns theory. In the first part of the essay I presented the classical origins, and in the second part I analyzed the evolution of the basic theory.

The idea of increasing returns is implicit in Adam's Smith's theorem on the effect of the division of labor and in Marshall's discussion of industry or sectorwide external economies. An important point in Marshall's theory is the fact that division of labor and mechanization create a tendency to increase the scale of production, and the chief advantages of production on large-scale are: economies of skill, economies of machinery and economies of materials. Besides, the aggregate scale of production results on growth of wealth, and an increase of this types of economies has a very important influence in the way of determining the supply price of a commodity. For Marshall, increasing returns could not coexist with a competitive market.

Young developed a dynamic model under conditions of increasing returns which stem from the division of labor along with roundabout methods of production. The indirect (roundabout) methods of production and the persisting search for markets by

modern industries, might have convinced Young that economic progress can best be understood by the Smithian theorem: the division of labor is limited by the extent of the market. The potential for expansion through specialization is unlimited and is possible even without population growth or new discoveries. However, this process of expansion is gradual for two reasons: learning of new skills and geographic adjustments by labor takes time, as does intertemporal capital accumulation. Young elaborates on the connection between the division of labor and increasing returns. The advantage of the division of labor among industries is secured by fuller realization of roundabout methods of production that become feasible and economical when their benefits can be spread over a large final output. This view of industries as an interrelated whole network is dramatically different from the neoclassical view of competitive firms under constant returns to scale.

Kaldor objected to the use of constant returns to scale as a device of economic analysis. As an alternative, he revived Allyn Young's ideas and applied his analysis to international trade. For Kaldor, specialization and then industrialization, gives rise to the effect of superior competitive power of the more efficient and dynamic countries, as compared to the others. To reduce, and maybe to avoid this effect of superior competitive power of the countries facing increasing returns technologies, it is necessary a carefully designed tariffs and subsidies system. This instrument is a necessary condition for industrialization to take place in small countries.

The theory of increasing returns finds its most widely recognized applications in the

new theories of endogenous growth, well represented by Arrow and Romer. Arrow's model of learning by doing shows that the learning effect in the economy is the cause of increasing returns, and the presence of increasing returns is possible even if we consider an economic environment of perfect competition. This is understandable if the distinction between private and social technology is made. Romer, who quite explicitly ties his work to Allyn Young's emphasis on specialization, identifies technology as the output that emerges from a production process that exhibits increasing returns to scale. Romer formulates Young's growth theory as an intertemporal optimization problem. He introduces an all-purpose capital good "Z", which is used to produce intermediate inputs and demonstrates that competitive equilibrium is suboptimal under the conditions of increasing returns due to specialization.

As said before, in the wide economic literature we can find a great number of studies involving the increasing returns issue, and the most relevant ones were presented here. However it is important to mention that, even when the work regarding increasing returns has been very extensive, there are still many puzzles. Nowadays for example, one of the main concerns between the macroeconomists has been how to incorporate the problem of nonconvexities (among which we can find the presence of increasing returns technologies) in business cycle theory. So the door of research in this issue is still open.

REFERENCES

- Arrow, Kenneth J, "The Economic Implications of Learning by Doing", *Review of Economic Studies*, June 1962, 39, 155-73.
- Buchanan, James M, and Yoon, Yong, J, *The returns to Increasing Returns*. The University of Michigan Press, 1993.
- Coase, R. "The Nature of the Firm", *Economica* 4, November 1937, 386-405.
- Davenport, H. J., *The economics of Alfred Marshall*. Reprints of Economics Classics, New York, 1965.
- Farmer, Roger E., "Macroeconomics", UCLA Lecture Notes, 1995.
- Leijonhufvud, Axel
- 1988. "Information Costs and the Division of Labor"
- 1986. "Capitalism and the Factory System", in R. Langlois (ed.), *Economics as a Process*. Cambridge University Press.
- "The individual, the market, and the industrial Division of Labor", Center of Computable Economics, UCLA, May 1993.
- Marshall, Alfred, "Principles of Economics", Macmillan, 9th Ed., 1961, Book IV.
- Romer, Paul M, "Growth based on Increasing Returns due to Specialization", *American Economic Review*, May 1987, 77, 56-62.
- Smith, Adam, *The Wealth of Nations*, The University of Chicago Press, 1976.
- Stigler, G. "The Division of Labor is Limited by The Extent of The Market", *Journal of Political Economy*, 59, June 1951.
- Targetti, F., and Thirlwall, A.P., *The Essential Kaldor*, Duckworth, 1989.
- Young, Allyn A, "Increasing Returns and Economic Progress", *The Economic Journal*, December 1928, 38, 527-542.