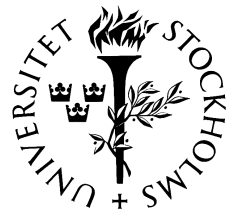
Seminar Paper No. 675

THE THEORY OF INTERREGIONAL EXCHANGE

by

Bertil Ohlin



INSTITUTE FOR INTERNATIONAL ECONOMIC STUDIES
Stockholm University

Seminar Paper No. 675

THE THEORY OF INTERREGIONAL EXCHANGE

by

Bertil Ohlin

translated by

Christina Lönnblad

Papers in the seminar series are also published on internet
in Adobe Acrobat (PDF) format.
Download from http://www.iies.su.se/

Seminar Papers are preliminary material circulated to
stimulate discussion and critical comment.

October 1999

Institute for International Economic Studies
S-106 91 Stockholm
Sweden

PREFACE

About the beginning of his work on international economics, Bertil Ohlin wrote in his memoirs: ”I started [in 1921] to write on the foundations of an approach to international trade theory that was to some extent new and for which I received the inspiration during a stroll on [the popular promenade] Unter den Linden in Berlin in 1920.”

Ohlin became a world famous economist on the strength of his monograph *Interregional and International Trade*, published by Harvard University Press in 1933. But before that, he wrote a dissertation for the degree of *licentiate*, finished in the spring of 1922, and a doctoral dissertation, published in 1924. Both were in Swedish.

The doctoral dissertation, entitled *The Theory of Trade*, was translated into English and published in 1991, together with a new and full translation of Eli Heckscher’s pathbreaking article *The Effect of Foreign Trade on the Distribution of Income*. The volume containing Ohlin’s dissertation and Heckscher’s article is called *Heckscher-Ohlin Trade Theory* and was translated, edited and introduced by M. June Flanders and myself.

The turn has now come to Ohlins *licentiate* dissertation. This work, the earliest by Ohlin on trade theory, deserves to be known to an English-speaking audience. It contains Ohlin’s short version of the Heckscher-Ohlin model, together with extensions and modifications that were to be dealt with more extensively in the doctoral dissertation, and even more in the monograph. The translation was made by Christina Lönnblad. She has been faithful to Ohlin’s somewhat involved style of writing.

Harry Flam

Table of contents

Introduction*

In most presentations of economic theory, the theory of international exchange, with its classical origins, stands out uniquely from the general theory of price formation, the origins of which are very different. This is most natural, since the former is based on the assumption that the values of goods are proportional to the quantity of labor used – an assumption which has been completely abandoned in the general theory of price formation in order to treat all factors of production equally.

This dualism constitutes an obstacle to constructing a consistent economic system. International exchange should not hold a unique position in such a system; it should be treated as a special case of general exchange, although certainly distinctive in some respects. It thus seems natural to build a general theory of exchange between different regions, based on the general principles of price formation. This theory will thus be valid for regions of any size; that is, for individuals as well as for regions and countries. The theory of international trade will then be a complementary study of exchange between a particular type of regions, that is, countries.

The present thesis is an attempt to solve the first part of the problem, that is, to present the general theory of interregional exchange. It is presented as a direct extension of the general theory of price formation, in accordance with the above views. The study of the latter part of the problem – that is, investigating the exchange between certain specific regions, above all, but not only, countries – has not even been carried out tentatively and, furthermore, it would require more space than has been considered appropriate in the present thesis.

* The translator thanks Harry Flam and June Flanders for useful comments on the translation.

Chapter I

The problem

The general theory of price formation shows how the prices of goods and factors of production are determined to create a balance between supply and demand. The factors of production are directed towards the production of the goods which, with the given income distribution, gives the maximum satisfaction of wants.[1] In this theory, where the fundamental nature of the causality is explained, goods as well as factors of production are assumed to be perfectly mobile. The question of the location of production is certainly of secondary importance and will thus not be dealt with here.

Under this assumption, the theory of exchange is relatively simple. An individual endowed with certain factors of production carries out the tasks for which he is best suited, that is, for which he receives the highest compensation, and exchanges them for the other tasks he requires with other individuals. The increase in productivity due to division of labor is allocated among the different individuals and constitutes the gains from division of labor and exchange.

As mentioned, price formation directs all factors of production to the utilization where they will command the highest price. It is thus obvious that one factor of production will command a higher price than another. Furthermore, there is nothing strange about the fact that although factor A would obtain a higher price in a certain employment than factor B, factor B is still the factor used. This is due to the fact that A commands an even higher price in another employment.

If the limited mobility of goods and factors of production is included in the argument, the question of interregional exchange arises, besides this purely individual exchange. The location of production is then no longer arbitrary and without effect on productivity, but is dependent on the local distribution of factors of production as well as the mobility and divisibility of these and the goods.[2] Since natural resources are

[1] Given wants must be assumed and the study deals with their maximum coverage. The hierarchy of wants is, from this point of view, given by the distribution of incomes and it is thus not the best hierarchy of wants from any point of view but a certain (in any way) fixed hierarchy of wants. As soon as there is also a choice between different hierarchies of wants, we enter the area of ethical values and lose the required fixed starting-points for the theoretical economy under discussion. The author's point of view is based on an unconscious hedonism which is, moreover, … and …. [undecipherable].

[2] This issue is treated more extensively in chapters 3-5.

immobile and other factors of production can only be moved from one place to another relatively slowly, the supply of factors of production remains different in different areas. Furthermore, due to the limited divisibility of the factors of production, the scale of production becomes an important factor in determining the production costs. The producing unit with the lowest costs requires production at a large scale and distribution of the products over a corresponding area. For these two reasons, division of labor will take place between different areas, since different industries are concentrated in different areas. The costs of transporting the goods tend to counteract this division of labor and, thus, also the ensuing exchange, but they cannot completely prevent it. Nevertheless, they do affect the division of labor and the terms of trade.

What are thus the effects of exchange between local units, that is, regions, resulting from the above-mentioned factors?

Exchange between individuals is either physical or legal. Individual exchange is the only kind of exchange. Individuals can, however, be grouped in many different ways, which makes it possible to study their exchange both with individuals *within* the same group and individuals from other groups. If the study is limited to the latter case, the different individuals in each group can be put in the background and each group can then be considered a complete unit in itself. Exchange between the groups will then take place.

Such groupings can be made in different ways. Within the theory of distribution, for example, three groups are formed, workers, landowners and capitalists, and the rules for the conditions for exchange between labor, land and capital are studied.

For these reasons, a local division and a study of the conditions for exchange between these locally limited groups, regions, are of particular interest. This can be done in different ways, however. Since the main aim is to modify the theory of price formation in general, due to the immobility of different factors, it seems desirable to make this study as general as possible and not limit it to certain types of regions. When the importance of the nature and effects of the factors studied has been established in the various cases, the general theory of interregional exchange is slightly adapted to the situation in each particular case.

The main reason for dividing a country into different regions is the uneven distribution of the factors of production. Certain factors of production are more abundant in one part of the country, others in another. The fact that this immobility remains is, in turn, due to the limited mobility of the factors of production. This immobility varies a great deal between different factors; from perfect immobility to almost perfect mobility. In general, however, limited mobility is sufficient for differences to remain unchanged, if the periods considered are not too long, although the mobility will vary with the selection of regions. For the basic treatment of price formation and exchange of goods, it is most convenient to begin by assuming that the factors of production cannot move from one region to another, that is, by assuming perfect interregional immobility[3] and then, if an adjustment seems necessary, adjust the results with regard to the effects of the more or less high degree of interregional mobility, which might exist in one case or the other.

Studying price formation and exchange in a country with such regions will thus, as indicated, be equal to determining the *causes* for locating production in one region or the other and the *effects* of this choice**.** When the limited mobility and divisibility of the factors of production, on the one hand, and the limited mobility[4] of the goods, on the other, are the final determining factors for location, as will be illustrated later, the pure theory of price formation is modified with regard to the effects of limited mobility and divisibility.

This modification will probably be gradual and, for this reason, the assumption that the goods and the factors of production are perfectly mobile is, at first, maintained. When the effects of the interregional immobility of the factors of production have been studied under this assumption, and the results have been modified with regard to the mobility of certain factors of production, the obstacles to the mobility of the goods are given due consideration. Finally, the impact of the limited divisibility of the factors of production on the price formation mechanism, that is, the increasing and decreasing returns, is studied.

[3] From the point of view of exchange, such a region will then, in certain ways, be equivalent to an individual in the pure theory of price formation. It has certain productive forces which will be attracted to the utilization where they fetch the highest price, without *leaving the region*. Nor can an individual's productive forces be moved outside himself. His tasks, however, that is, the results of his productive forces, can be transported, just as the goods of a region can be sent to other regions. An important difference, however, is that the individual, but not the region, can be moved in its entity.

[4] Expressed as difficulties in transportation, particularly transport costs.

Chapter 2

The theory of interregional exchange under simplified assumptions

A. Conditions for interregional exchange of goods

According to the outline presented in the previous chapter, the aim is now to study the price formation mechanism in an isolated area consisting of regions between which the factors of production are immobile. There are two different ways in which to proceed. One is to begin with price formation in an area with perfect mobility and analyze the changes caused by the fact that factors of production cannot be transferred from one region to another. The other is to begin with two isolated areas with perfect mobility within each area and analyze the consequences of contacts between these areas when goods, but not factors of production, can move freely from one area to the other. In both cases, the pure theory of price formation is assumed and the desired modification of this theory is achieved. The latter method is preferable, however, for a detailed study of the exchange of goods and its preconditions and effects. Under certain circumstances, it is the starting point of the exchange of goods that causes certain changes and for this reason, the study *immediately* supplies the desired information.

Two isolated areas, between which autarky is removed for some reason, constitute the starting point of the following analysis. These two areas are assumed to have their own independent monetary systems.[5]

When a connection has been established, there exists a certain exchange rate between the two currencies.

This exchange rate must be such that, if exchange of goods takes place, the imports of each region must exactly equal its exports. If the imports of either of the regions exceed its exports, the former can then not be entirely covered by the claims of the latter; the demand for “foreign” currency thus exceeds the supply and the value of the foreign currency increases. Exports are thus stimulated and imports are discouraged, until there is a balance in the exchange of goods. When this balance and a fixed exchange rate between the currencies have been established, all prices of goods in both areas will be equal.[6] The types of goods that would have commanded a

[5] Two isolated areas can, naturally, not be assumed to have a common monetary system. [Sic].
[6] The condition is the absence of transport costs for the goods.

higher price if produced “domestically” are imported while, once more, the types of goods that would have commanded a higher price if produced in the other region are exported to that region.

Exchange of goods can only take place due to such a difference in prices. When this difference has been removed through exchange, no *new* exchange will occur, while the ongoing exchange still continues to the extent required for price equalization.

At the moment before the connection is established, there is no question of a difference in the *absolute* prices of goods, since each region has its own monetary unit with no exchange rate for the monetary unit in the other region. The condition for exchange of goods to occur will then be a difference in the *relative* prices of goods. This is not only a *necessary*, but also a *sufficient*, condition. If the relative prices are different, all prices of goods will be lower in region A than in region B or all prices lower in region B than in A or some prices lower in A and some in B, notwithstanding the exchange rate which was at first assumed between the currencies. In the first case, the goods will flow from A to B, without B being able to pay for them by exports to A. In other words, there will be a demand for A’s currency in B, in order to import goods from A. In A, there is no demand for B’s goods, however, and accordingly, there will be no supply of A’s currency. Under these circumstances, the value of A’s currency must increase in relation to B’s currency. Thus, all prices of goods from A will increase, *expressed in B’s currency* and the prices of goods from B will decrease, *expressed in A’s currency*. It will thus be profitable to import certain goods from B to A, since only some goods, and not all as in the previous examples, can be imported from A to B. Equilibrium occurs when the exchange rate between the two currencies has reached the point where imports and exports are equal.

The effects would be analogous if all prices of goods were lower in B than in A. In the third case, when some prices are lower in A and some in B, imports as well as exports will occur, according to the exchange rate established at the starting point. If A’s imports exceed its exports, the demand for B’s currency will exceed the supply and its price will increase. The same applies if B’s imports exceed its exports. Equilibrium will occur only when an exchange rate has been established, thereby establishing an equilibrium between imports and exports.

A difference in relative prices thus necessarily entails a difference in absolute prices, so that some prices will be lower in A and some in B, and exchange of goods will take place.

Using the term production costs instead of prices – which equal the costs, since these cover remuneration to all factors of production – this can be expressed as follows: the effects of a difference in comparative costs are that some goods can be produced at a lower cost in one region than in the other, while the opposite applies to other goods. Thus, exchange will occur and each region can concentrate its production on the goods it can produce at the lowest cost.

It now remains to be shown that exchange of goods cannot occur, if comparative costs and prices in both isolated regions are equal. Whatever exchange rate is assumed between the two independent currencies, *all* costs and prices of goods must be lower in A than in B or lower in B than in A, or equal in A and B. In the former case, the value of the currency in A increases in relation to the currency in B; in the latter case, the currency in B increases in relation to the currency in A. The exchange rate reaches the point where all costs and prices of goods in A equal all prices of goods in B, without exchange of goods having occurred. According to the assumption of equality of comparative costs, some costs cannot be lower in A and some in B. *Different comparative costs thus constitute the necessary and sufficient condition*[7] *for interregional exchange.*

It then follows that one region cannot be competitively superior to another in the production of *all* goods. A difference in comparative costs creates an exchange rate between the monetary units in the two regions, which makes the costs for some goods higher in one region and for others in the other region. A completely different issue is that using other terms of costs as a yardstick, for example the amount of labor used,[8] might yield the result that a country is superior in all kinds of production, but that exchange of goods is still profitable. The paradox of this reasoning is removed if one considers that the yardstick of exchange is not the quantity of labor put into a good but rather the price of the good expressed in terms of money, where the price

[7] The condition for the entire chapter II is that the factors of production are perfectly divisible, that there are thus no increasing returns and that the goods are perfectly mobile.
[8] Or similar units of quantity for productive forces.

expressed in the currency of one country is converted into the currency of the other country according to the exchange rate.

Take, for example, Ricardo's classical example: Portugal can produce one unit of wine in 80 working days and one unit of cloth in 90 working days. The corresponding figures for England are 120 and 100. Portugal can thus produce both types of goods "at a lower cost" than England. Nevertheless, Ricardo shows that the production of wine in Portugal and of cloth in England will increase total production and, thus, be profitable.

Note that if Portugal can produce these types of goods with less labor than England, this is either due to the fact that labor in Portugal is of superior quality or that there is a relatively more abundant supply of other factors of production, for example a favorable climate and fertile land, which can partly substitute for labor. In both cases, the marginal productivity of labor can be considered as higher in Portugal, which means that wages are higher in Portugal,[9] since Ricardo does not consider compensation to other factors of production or considers it to be proportional to compensation to labor. The number of working days is thus not a way of expressing production costs and, accordingly, not a way of expressing competitiveness either. The main point is thus that the higher wage level in Portugal makes it more expensive to produce cloth there than in England.

There must thus be a difference in relative costs and prices of goods, *expressed in terms of money*, for exchange to occur. This might also occur if the relative costs, estimated on basis of quantity of labor, are equal. Assume, for example, that two types of goods require an equal amount of labor, but that the one good requires much capital and little land, while the other requires little capital and much land. If capital rent is low in region A and high in B, while the opposite applies to land rent, the costs will be lower in A for the first and lower in B for the second good, and exchange of goods will occur.

This leads to the question of the reason for the differences in comparative costs and prices of goods; an issue hardly dealt with in most studies.[10]

[9] Higher real wages in Portugal and equal prices of goods in both countries must mean higher wages in Portugal.
[10] See Heckscher: The effect of foreign trade on the distribution of income. *Ekonomisk Tidskrift* 1919.

The costs of producing a good depend both on the quantities required of the different types of factors of production and their prices. What is, then, the meaning of a difference in relative costs, with regard to these two factors?

The answer to this question is probably most easily found by a mathematical formulation. The quantities of different factors of production required for producing a good are denoted by a_{ij} in region A and α_{ij} in region B. The prices of the factors of production are denoted by q_1 and g_i, and the prices of the goods by p_i and v_i, respectively.[11] The following equations then show that the price of each good in each country equals the production cost of this type of good.[**]

	A	B
(1)	$a_{11}q_1 + a_{12}q_2 + \ldots + a_{1r}q_r = p_i;$	$a_{11}g_1 + \alpha_{12}g_2 + \ldots + \alpha_{1r}g_r = v_1;$
	$a_{21}q_1 + a_{22}q_2 + \ldots + a_{2r}q_r = p_2;$	$\alpha_{21}g_1 + \alpha_{22}g_2 + \ldots + \alpha_{2r}g_r = v_2;$
	.....	.....
	$a_{n1}q_1 + a_{n2}q_2 + \ldots + a_{nr}q_r = p_n;$	$\alpha_{n1}g_1 + \alpha_{n2}g_2 + \ldots + \alpha_{nr}g_r = v_n$

If autarky is removed, exchange will occur, provided that the relative prices of the goods are not equal, that is, provided that the following condition does not hold:

(2) $p_1 : p_2 : \ldots : p_n = v_1 : v_2 : \ldots : v_n .$

This will be the case, however, and exchange can thus not occur, if

(3) $q_1 : q_2 : \ldots : q_r = g_1 : g_2 : \ldots : g_r ;$

that is, if the relative prices of the factors of production are equal, for if condition (3) is fulfilled, $a_{ij}=\alpha_{ij}(i=1,2\ldots n;j=1.2\ldots r)$ holds. The different a and α, that is, the technical coefficients, give the possible proportions in combinations of the factors of production; they are thus both dependent on certain technical conditions and the relative prices of the factors of production. If the latter are equal in two regions, the technical coefficients must be equal, since the technical conditions, which are related

[11] Cf. Cassel: *Theoretische Sozialökonomie.*

[**] Translator's note: Notation is as in the original, including subscripts.

to the technical characteristics of the goods and the factors of production, are equal everywhere.[12]

Condition (3) can also be written:

(4) $q_1 = lg_1; q_2 = lg_2; .....q_r = lg_r;$

where *1* is an arbitrarily positive quantity. When $a_{ij} = \alpha_{ij}$:

(5) $p_1 = lv_1; p_2 = lv_2; ..... p_n = lv_n;$

Since (5) is the same condition as (2), it appears that equality between the relative prices of the factors of production means equality between the relative prices of the goods and thus, no exchange will occur.

When a connection has been established between the two regions, the absolute prices of goods must be equal. The exchange rate will thus be *l*.

Condition (2), that the relative prices of the goods should be equal, can be fulfilled in another way than by (3), that is, if the following conditions are satisfied:

(6)
$$\begin{aligned}
&a_{11} : a_{12} : ..... : a_{1r} = \\
&= a_{21} : a_{22} : ..... : a_{2r} = \\
&= ..... = \\
&= a_{n1} : a_{n2} : ..... : a_{nr} = \\
&= \alpha_{11} : \alpha_{12} : ..... : \alpha_{1r} = \\
&= \alpha_{21} : \alpha_{22} : ..... : \alpha_{2r} = \\
&= ..... = \\
&= \alpha_{n1} : \alpha_{n2} : ..... : \alpha_{nr};
\end{aligned}$$

If all types of goods contain the same proportions of all factors of production in *both* regions, exchange cannot take place, since the relative prices will then be equal.

It turns out, however, that (6) is not a new condition for failure of exchange to take place; it cannot be satisfied unless (3) is fulfilled. If the relative prices of the

[12] See appendix 3 [appendix is missing].

factors of production in both regions are different, they cannot possibly be combined in the same way in production in both regions. The factors of production that are relatively scarce in A will thus be used sparingly in A and be replaced by the less scarce ones to the largest possible extent. The opposite applies to B. Such possibilities for substitution might always exist, although to various extents. Only if the relative prices of the factors of production are equal in both regions, can (6) be fulfilled. But then this condition will not be necessary. As shown above, the relative prices of the goods will then be equal, notwithstanding the proportions of the different factors of production in different types of goods.

There are, however, several ways of fulfilling (2). The different a and q, α and g can, of course, take on such values that (2) is fulfilled "by accident", without their assuming any *regular* relationship with each other.

Condition (3), that is, that the relative prices of the factors of production should be equal in both regions, is therefore a *sufficient but not a necessary condition* for equalization between the relative costs and thus also for the non-occurrence of exchange.[13] Reversed, we get the following phrase: *a difference in the relative prices of the factors of production constitute a necessary, but not a sufficient, condition for interregional exchange.*

In accordance with the classical procedure, two isolated regions have, so far, constituted the staring point. Being isolated, it must be assumed that each region has its own individual monetary unit. It is clear, however, that the result will basically be the same if two regions with a common monetary system, for example two parts of the same country, are considered. Under what circumstances might exchange of goods between two such regions not occur? Only if the *absolute* prices of goods are equal without exchange, that is, if the production costs for all goods are equal in both

[13] In the paper referred to, Heckscher has come to the conclusion that exchange of goods will not occur if all goods use the same proportions of factors of production in both regions. From the above, this is wrong, for it must apply to all goods in *both* regions. See equation (6). Due to this mistake, the fact that the condition "different relative scarcities of factors of production" in itself contains the latter condition concerning the proportions of the factors of production in the goods, has been overlooked. Moreover, the mistake seems quite fatal, since H; like Bastable, Pierson and others uses the unit productive force as a yardstick, when first studying foreign trade. This means that the factors of production are assumed to constitute fixed proportions of all goods in each country, although not equal proportions in *both* countries. Foreign trade is thus studied under an assumption which, according to H. himself, in itself precludes all kinds of foreign trade.

regions. If exchange of goods takes place, however, this shows[14] that the costs are lower in one region or the other as would have been the price of the goods, had no exchange occurred. The case when a difference in absolute costs also means a difference in relative costs, only differs from the case just studied in that a difference in relative costs due to the common monetary unit is *directly* expressed by some costs being lower in A and others in B, while in the previous case, this difference is only expressed by the exchange rate. There is thus only a formal difference in the conditions for exchange of goods, which means that the previous wording of the condition for interregional exchange, that is, different comparative costs, also applies to this case, and the *meaning of this condition* remains unchanged.

B. The mechanisms and effects of interregional exchange of goods

If the above holds, this means that the difference in the relative scarcity of factors of production is the cause[15] of the difference in comparative costs and thereby of interregional exchange, although the comparative costs might remain the same, even when differences of the first kind exist. In practice, it might reasonably be assumed that the relative scarcity of factors of production leads to differences in comparative costs and thus, to exchange of goods. In practice, the relative prices of the factors of production are thus not only a necessary, but also a *sufficient*, condition for interregional exchange.[16]

As has already been shown, the principle of this exchange is that each region produces and exports the goods it can produce at a lower cost than others, and imports the other goods. The costs of production and thus, the *absolute* prices, will be decisive. Under the influence of the forces creating an equilibrium in trade, prices are controlled in such a way that exports will take place and cover the costs of imports, as has been shown in the previous section.

The reason why the production costs for certain goods are lower in one region than in another is part of the above mentioned condition for interregional exchange, that is, the difference in the relative scarcity of factors of production. Since there is a

[14] Under the assumptions of chapter II.
[15] By 'cause' is meant necessary *and sufficient* condition. The wording of the text should not be a source of misunderstandings.

relatively good supply of certain factors of production in A and their prices thus are low, the goods requiring a relatively large share of these factors of production, can be produced at a low cost in that region compared to the types of goods requiring a relatively larger share of the more scarce factors of production. In B, where the former factors of production are more scarce and the latter less scarce than in A, the relation between the costs of different goods will be completely different. A will produce the goods for which it has a relatively large supply of suitable factors of production; and so will B.

A sparsely populated region, with much fertile land and a favorable climate, will export agricultural products and import industrial products from a region with a relatively rich supply of capital and technical skills.

Exports will then mainly consist of goods where the abundant factors of production are used in large quantities and the other factors in small quantities only, while imports consist of the types of goods requiring large quantities of the latter factors of production and factors of production not available in the region at all. The exchange of goods is thus an exchange of factors of production, in the sense that goods "containing" the relatively scarce factors of production are imported and goods "containing" less scarce factors of production are exported. It might be said that less scarce factors of production are exchanged for more scarce ones. The result is thus that the scarcity of factors of production is generally equalized.

The factors of production are assumed to be immobile between regions, and no transmission and direct equalization of the scarcity can thus occur. The mobility of the goods will thus, to a certain extent, replace the mobility of the factors of production, since exchange of goods leads to an equalization of the scarcity of factors of production, from the point of view of all traders.

There is thus a tendency that all regions get exactly the relative scarcity that would have existed, had there been no obstacle to the mobility of factors of production. *In other words, interregional exchange tends to create a uniform price structure of the factors of production.*

[16] When the transport costs of the goods and the increasing returns are not considered.

As shown above, different relative prices of the factors of production are a necessary condition for different relative prices of goods before autarky is eliminated, and thus also for interregional exchange. Since exchange of goods tends to equalize these differences, it also tends to remove its own reason of existence. It must not be assumed, however, that exchange of goods goes on for a period of time, until equalization occurs, and then ceases altogether. The previous difference in the relative scarcity of factors of production would then almost immediately recur. The result is instead that exchange of goods will take place to the extent required for equalizing the relative scarcity of factors of production, and then *remain the same*. No *new* exchange of goods will occur after the equalization – since there are no conditions for such an exchange – but the existing exchange will persist.

If goods as well as factors of production were perfectly mobile, the entire world would constitute an economic unit in the real sense of the word and the total supply of different factors of production would determine their relative scarcity. As mentioned, a free exchange of goods tends to give the same result and might also do so, under certain conditions. What are these conditions, or conversely, what generally prevents this tendency from being fully realized?

It is difficult to find a general answer to this question. On the most profound level, there is probably a certain disproportion in the distribution of the factors of production and their use in production and thus, the same results cannot be obtained as if the factors of production were mobile, notwithstanding the location of production. The exchange of goods can only create a situation where one type of good will be produced here and another there, that is, that each good is produced in that or those places where the most favorable of the existent combinations of factors of production is to be found. Had the factors of production been mobile, however, an even more favorable combination could have been created in many cases, due consideration given to the worldwide scarcity of these factors of production. Due to the exchange of goods, the utilization of the factors of production which would have occurred at perfect mobility is not achieved.[17] A certain factor of production will be more scarce in one region, and less scarce in another.

[17] For further discussion of this issue, see appendix 4 [appendix is missing].

Assume that the production of a good, wine for example, requires land[18] of a certain quality as well as a certain labor skill. A region fulfilling the condition of labor skill might only have a limited supply of land of the desired quality, while another region, which lacks suitable labor, has an abundant supply of such land. The entire world demand for that good must then be produced in the first region. This might only be possible through an intensive, capital and labor squandering, cultivation of land. Land will be extremely scarce and will fetch a high land rent, while land in the other region will be of insignificant or zero economic value, depending on how it can be used for other purposes.[19]

If labor had been mobile, all land of the desired quality would have been used, and thus, capital and labor would have been saved and land rent would have been equalized in both regions

When the supply of the different factors of production is such that all factors of production would not be fully utilized if prices were uniform, notwithstanding how production is distributed between different regions, there will be a necessary shift in prices. The prices of the factors of production, which would, to a certain extent, have been unemployed in A, will fall and thus, they will be utilized to a greater extent, while the prices of other factors of production, which would have been unemployed in B, will fall in B. The relative prices in A and B will thus differ.[20] This phenomenon of disproportion counteracts, but cannot entirely remove, the tendency of the exchange of goods to equalize the relative scarcity of factors of production, that is, to create uniform prices of the factors of production, unaffected by local distribution. This tendency results from the economy's "tendency towards maximum satisfaction". Free competition within an isolated region tends to produce the production patterns and the consumption best suited to the nature and intensity of the wants, by leading to uniform prices of factors of production and goods, and it also tends to have the same effect in a number of regions mutually exchanging goods. In the latter case, this tendency is counteracted by the above-mentioned phenomenon of disproportion, however. The smaller its effect, that is, the more uniform are prices, the more complete is the

[18] The term "land" here covers general natural conditions, such as climate.
[19] Cf. Heckscher.
[20] This question is discussed in more detail in appendix 4 [appendix is missing].

adaptation to wants and the larger is the extent to which the tendency to maximum fulfillment of wants is realized. Accordingly, the profits of the traders are determined by the extent to which a free exchange of goods can create an equilibrium between prices of goods in different regions and the extent to which it can reduce, if not entirely remove, the differences in the prices of the factors of production.

The difference in the relative scarcity of factors of production in different regions creates an uneconomic production pattern. In a certain region, a certain factor of production should be used very sparingly and everything be done to replace it with other factors of production, which are less scarce in that region. In another region, the latter factors might be relatively scarce and can only be utilized very sparingly. Certain factors might be abundant in one region and thus be utilized as such, while the other region might lack these very factors, and the other way round.

If there is a rich supply of labor in one region, A, and its price is thus relatively low, and land rent is relatively high, while there is a surplus of land and land rent is thus low and wages high in another region, B, the marginal productivity of labor must be relatively low in A but high in B, while the opposite applies to the marginal productivity of land. Wages and land rent thus equal the marginal productivity of labor and land. If part of the labor could be transferred from A to B, the production in the former region would obviously be subject to a relatively small decrease in relation to the increase in the latter. The result would be high marginal productivity instead of a low. Wages would increase in A and decrease in B, thus tending towards the same level, and land rent would fall in A and increase in B, and thus also tend towards the same level in both regions. An equalization of the relative prices of land and labor would then result. As long as this equalization is not complete, there might still be a profit, but, with complete equalization, it will have reached its maximum, that is, production will have been adjusted to wants, as far as is possible with an uneven distribution of income.

If the factors of production are immobile, and labor can thus not be transferred from A to B, the result will still be the same, provided that the requisite equalization of the relative prices of the factors of production can be obtained by exchange of goods. This is the case, at least to a certain extent, since exchange of goods means an

exchange of less scarce factors of production for more scarce ones, in both regions. The result will thus be the same in both cases: since the relative prices of the factors of production have been equalized in both regions and have reached the position determined by world supply and world demand, their combination will be the same as it would have been, had the factors of production been mobile. The only difference is that, in the latter case, it would have been equally advantageous to produce any type of goods anywhere, while the location of production is now determined by the location of the factors of production.

The more uniform is the price structure of the factors of production, the higher is the extent to which "the tendency towards maximum satisfaction" is realized through international trade.

Under these circumstances, the profits from international trade cannot be exactly estimated. The value of the total production in each country before and after the exchange of goods has been established could, of course, be statistically determined and the figures obtained reduced according to the change in the value of the currency, as expressed in price index calculations. Changes in production and consumption due to changes in the relative prices of the goods and, possibly, the supply of entirely new goods make the inherent logical weakness, or insustainability, of the calculations of the price index particularly obvious in this case. It is thus likely that such estimates would not be a good illustration of the actual profit.

There is even less likelihood of obtaining the desired result by using a central concept of classical theory, that is, the terms of trade, thereby meaning the terms of trade between domestic and foreign goods, which is further illustrated by Bastable's example.[21] One unit of "productive power" in country A can produce 10 units of good *x* or 20 units of good *y*, while one unit in B can produce 10 *x* or 15 *y*. According to the law of comparative costs, A will only produce good *y* and B good *x*. What are then the terms of trade between *x* and *y*? These will be somewhere between 10 *x* =15*y* and 10 *x*=20*y*, that is, the limits given by the comparative costs; to be precise, at the point where A's demand for *x* exactly equals B's demand for *y* and the balance of

[21] *Theory of International Trade*. London 1903.

trade is in equilibrium. Specialization increases total production from $20x + 30y$ to $20x + 35y$ and the profit is thus $5\ y$.

There is a major weakness in this reasoning about terms of trade, however. It is not suitable if, as required, the fact that exports as well as imports consist of a large number of different types of goods is taken into account. In analogy to price indexes, one might, of course, create quantity indexes and thus obtain a figure for the quantity of exports and another for the quantity of imports. A comparison between these figures would thus show the terms of trade in interregional trade, which must be considered separately from the terms of trade which would have existed without trade. The same difficulties as when constructing a price index would then appear, however; prices as well as quantities of goods would have changed and entirely new types of goods might also have been included. But there is also another difficulty, which should be mentioned already at this stage, that is, that not all types of goods will be subject to interregional trade when the assumption of the absence of transport costs and other such costs is removed. The effect of interregional trade on their production and on price formation will then not affect the terms of trade. Under these circumstances, this condition does not seem at all suitable for shedding light on the question of the size of profits. The general reasoning about uniform prices and "the tendency towards maximum satisfaction" will have to suffice.

If the size of profits cannot be estimated, it is, obviously, just as impossible to determine how the profits are to be divided between the traders. It is as impossible to determine the proportions in one case as in the other. It is, however, worth studying in what *direction*, advantageous or disadvantageous, changes in the conditions for trade affect different regions. The study of these questions should be postponed, however, and be made in a more general context, investigating the effects in different cases, once the simplifying assumptions regarding the mobility of the factors of production, transport costs and other such factors have been removed. All problems, for the solution of which abstracting conditions are not absolutely necessary, should be studied under conditions as close to reality as possible, in all respects.

In summary, this part of the study concerning the conditions, mechanisms and effects of interregional exchange, gives the following results:

Conditions: It can be derived from the law of comparative costs that the relative prices of the factors of production being different in both regions is a necessary, and in practice, also a sufficient, condition for interregional exchange.

Mechanisms and effects: Interregional exchange tends to equalize the relative scarcity of factors of production, that is, achieve a uniform price structure of the factors of production. Thus, it tends to remove its own reason for existence. This tendency is usually not fully realized, however, since these factors are disproportionate. When the tendency has been realized, to the extent permitted by the obstacles, trade remains unchanged but *no* new trade occurs.

The tendency towards a uniform price structure means a tendency towards maximum satisfaction of preferences.

A general comparison between the price formation mechanism in two isolated regions and two regions exchanging goods will probably give a more general picture of the causality of interregional exchange. Due to the exchange, these two mechanisms become one only, where each factor is connected to all other factors in both regions. Any kind of change in the economy of one region can thus affect its exchange with the other region. Due to the mutual interdependence between all factors, the effects will spread to all parts of the price formation mechanism. Keeping this context in mind is important for studying several of the theories of international trade, for example when assessing the attempts to give the goods which are actually part of the exchange a unique position in determining the equilibrium for foreign trade and the exchange rates.

This mutual interdependence and the effects of exchange of goods on the price formation mechanism in two regions previously isolated from each other, can be most simply and most easily illustrated by equations.[22]

So far, the reasoning has been based on the assumption that there can exist two regions only. It is, however, clear that the conditions for interregional exchange as well as its mechanisms and effects will remain exactly the same if an arbitrary number of regions are considered. No recapitulation of the reasoning should be required under

this assumption. It should only be noted that, if the factors of production in two isolated regions, A and B, command the same relative prices, while their prices differ in region C, exchange will take place between A and C and B and C, when a connection has been established. Thus, the scarcity of factors of production in both A and B is subject to a change, as shown above. It might thus be the case that the change in A differs from the one in B, that the relative prices of the factors of production in A and B will differ, and that exchange between these will also occur.

Another simplifying assumption, made above, is that the factors of production are completely mobile within each region. Naturally, this is an abstraction. Land and natural objects are immobile and other factors of production are not even perfectly mobile within fairly small regions. The location of production *within* regions is thus a problem of the same kind as the distribution between regions. This is natural since the division into regions is entirely random and the present problem concerns the general location of production and exchange. The abstraction makes the essential feature of exchange seem far more simple than would otherwise have been possible. Furthermore, it is not difficult to abandon the arbitrary division into regions after having studied the limited mobility of goods and the limited divisibility of factors of production, and give a general picture of the location of production and exchange without abstract assumptions. This abstraction entails another advantage, however, besides serving as a bridge to presenting reality in its entire complex context. When the general results reached by the study of interregional exchange can be applied to trade between specific regions, above all, countries, the location of production within the country is of secondary importance. The distribution of production between countries and the exchange between these will then be the most interesting question. Thus, the assumption of perfect mobility within the country is the most rewarding one, since it best illustrates the main questions.

For these reasons, the assumption of perfect mobility of factors of production *within* regions is maintained for the time being. Only when it is time to present a general picture of the location of production and exchange, as it appears at the final

[22] See appendices 3 and 4 [appendices missing].

stage of the general theory of interregional exchange, will this assumption be removed.

The third simplified assumption made above is the interregional immobility of the factors of production. For several of these factors, such as labor and capital, this immobility is far from perfect, however. It must therefore be shown to what extent the interregional mobility of the factors of production affects the results obtained under the assumption of immobility.

Chapter III

Modifications due to the interregional mobility of factors of production

To investigate the general causality, it is convenient to start by considering a smaller or larger number of variable factors as constant. On this basis, create an equilibrium and investigate the effects of variations in these factors on the nature of the causality and then *modify* the earlier presentation with due consideration to these variations. The presentation can then be completed with an illustration of the *shift*[23] in this modified system through changes in the determining factors.

This method which is, for example, used to a great extent within mechanics and astronomy, was used in economics when constructing the general theory of price formation. It is, however, quite common not to proceed further than the first part of the study and be satisfied with a static theory. The same method has been used in this study. The assumption of the interregional immobility of factors of production does mean that the supply of factors of production is assumed to be constant. The extent to which the nature of the causality is subject to change when certain factors of production are actually transferred from one region to another then remains to be studied. The third step, that is, illustrating the shift in the entire system, will only be taken at a later stage.

Consider two regions where the relative scarcity of factors of production would be very different, if there were no exchange of goods. This difference is reduced by the exchange of goods, but cannot be entirely removed. If the factors of

[23] which, accordingly, does not affect the nature of the causality.

production are partly mobile and drawn to the location where they will get the highest compensation, some of the relatively abundant factors in A will eventually move to B and the other way round. There is, thus, a certain tendency to equalize the relative scarcity. A region with a high interest rate and low wages exports labor to and imports capital from a region with a low interest rate and high wages. There is thus, eventually, a change in the conditions for interregional exchange, its extent and the entire price formation mechanism. Above, this change was called the shift in the system.

A completely different question is whether the ongoing exchange at a certain point in time will be different than it would have been with interregionally immobile factors of production. In other words, will their mobility affect the nature of the causality? Will production and exchange at a certain point in time be affected by the fact that a certain inflow and outflow of factors of production is possible?

The reply to this question naturally depends on whether people consider present or potential changes in the relative scarcity of factors of production in their economic behavior. If the mobility of labor is first considered – land and natural resources are, of course, perfectly immobile – it will probably be found that it has a very small effect on interregional exchange. When determining the sales price of his goods, a producer considers the prices of the different factors of production, including labor, as they are at present, and hardly considers future changes in wages, in relation to the prices of other factors of production, which might be incurred by ongoing immigration. Ongoing immigration has no considerable effect on production and price formation, at present, although it causes a gradual and, in the long run, perhaps considerable, shift. The nature of the causality will thus, to a large extent, be the same as explained under the assumption of immobile factors of production. The *current*, and not the future, distribution of labor constitutes part of the determinants of prices.

The above applies to labor in general, without considering its various qualities. The impact of mobility becomes more considerable, if the large differences in the qualifications of labor are considered. A country with an abundant supply of unqualified labor can have a scarcity of more qualified kinds of labor, that is, people with technical and organizational skills. Even a relatively small immigration of such

individuals can have considerable effects on the structure of the economy. Numerous examples can be found in certain parts of Africa, where the colored population dominates in numbers but would be unable to run its present industries without its white leaders.

Thus, the mobility of labor is often of major importance for the economy of a country, mainly due to the possibility of obtaining certain absent qualities. The conclusion to the above reasoning on the effects of the general mobility of labor remains more or less the same. There is a shift in the equilibrium, but hardly a more considerable change in the nature of the causality. The imports of a country depend on the extent to which the demand for goods is actually satisfied by domestic production and not on the prospects of one type of good or another being produced if the missing labor were imported. It might, of course, be assumed that importers would decrease their stocks, if they knew that a factory to be run by foreign engineers and foremen was under construction. A decrease in their demand might then tend to reduce the price abroad, which means that prices will be affected not only by the current, but also by the future distribution of labor. In general, such effects should be quite insignificant. Only after a certain period of time will the immigration of labor have a more considerable effect on production. It is probably very rare to discount the effects of immigration *before* immigration has taken place. Price formation is mainly determined by the actual supply of labor, and not by expected changes.

It might thus be assumed that labor mobility does create a gradual shift in the entire equilibrium, but no considerable change in the nature of the causality, which, on the contrary, is more or less as illustrated under the assumption of interregional immobility.[***] In a way, the same is true of capital. On the one hand, this is due to the fact that capital rent is often approximately the same in all parts of a region with a common monetary and banking system, and for this reason, no major transfer from one part to another occurs. On the other hand, the obstacles to the mobility of capital between different countries are so high that the transfer of capital cannot considerably affect the capital resources of a country in the short run, despite the different interest rate levels. There is thus a gradual shift in the equilibrium, but no change in the nature

[***] Translators note: “mobility” in the original.

of the causality. *From one point of view*, it might thus generally be said that the mobility of capital and the mobility of labor do not considerably modify the above presentation of the causality of interregional exchange.

This reasoning is incomplete, however, since one important factor is ignored. Capital movements hold a unique position in relation to the movements of other factors of production, in the sense that they affect the "balance of payments" and thus, in a unique way, the entire equilibrium. *The condition that imports and exports of goods should be in equilibrium, no longer applies.* A relatively insignificant capital movement in relation to total capital resources will thus considerably affect the equilibrium. A country with continuous capital inflow imports more goods than it exports. The foreign goods can not only be paid for with exports of domestic goods, but also with borrowed money. The prices of goods in the country will then, to a large extent, depend on capital imports. If no such imports exist, prices must be such that imports entirely balance exports. Borrowing abroad creates another equilibrium, where a larger or smaller number of the domestic goods command higher prices, thereby decreasing the competitiveness with foreign countries. Thus, imports will increase and exports decrease. At this new equilibrium, the deficit in the trade balance will be exactly covered by capital imports.

The change in the prices of goods is only one side of the shift in the general equilibrium caused by borrowing. The prices and the utilization of the factors of production are also affected. Certain industries decrease, others increase; changes in income and the prices of goods will also lead to a restructuring of consumption. What has so far been said about the mobility of the factors of production has not required any substantial change in the presentation of the construction of the price formation mechanism, but in this case, the situation is different. Capital movements constitute a factor in the trade balance. The causality is thus entirely different than it would have been, had capital movements not been possible.

So much for the nature of the causality. The question of a shift in the equilibrium will only be studied later, in a different context. It should be observed, however, that at the same rate as capital movements change, there is a corresponding shift in the entire equilibrium. Since borrowings can change considerably and very

often, these shifts might not only be numerous but also significant, which constitutes a difference from the cases studied so far. Neither the mobility of labor nor that of the organizational factors will produce anything but slow shifts, and this also applies to the effect of capital movements, through the supply of factors of production.

The considerable effect of capital movements is due to the fact that they might be large relative to the imports and exports of a region. Very considerable changes in the equilibrium would thus be required to create an equilibrium in the entire balance of payments, in the absence of capital imports or exports.

Summarizing the study of the effects of the interregional mobility of the factors of production on the price formation mechanism, it appears that, due to the mobility of capital, a new factor is introduced into the balance of payments, which will determine the equilibrium together with the factors previously studied. However, the effects of mobility on the supply of different factors of production and thus, on the production pattern do create a continuous shift in the mechanism, but no important changes in the way it works.

Chapter IV

Modifications due to the limited mobility of goods

The pure theory of price formation builds on the assumption of perfect mobility, not only of factors of production but also of goods. Having studied the mobility of the former in the two previous chapters and having paved the way for the removal of all simplifying assumptions on this issue, the next step is to modify the presentation by considering the consequences of the obstacles to the mobility of the goods. Transport

costs are among the most important of these obstacles. Their importance varies with the value of the goods per unit of weight and volume. Items such as fine instruments, watch-springs and jewelry can be transported around the world for less than one per cent of their market value, while transport costs for heavy and bulky articles, such as bricks, come close to or, at times, even exceed production costs, even when transported fairly short distances. Most types of goods fall in between these two extreme cases.

Besides transport costs, different kinds of trade costs[24] also constitute barriers to the mobility of goods. With a common name, these could be denoted as the mobility costs of the goods. Due to these costs, one can no longer assume a single price for each type of good, but different prices depending on the distance between the place of consumption and the place of production. A difference in absolute costs[25] no longer suffices to induce an exchange between two areas. A good might be produced at a lower cost in one of these areas than in the other; if the difference is less than the mobility costs, the other area will still produce the article in question itself. No exchange will then occur.

Like the prices of goods, these mobility costs are variables determined by prices. They mainly consist of such costs as wages, costs of coal and costs of constructing means of transport and can be broken down into their components, that is, the quantity of each factor of production. In the price formation system, they hold a position completely analogous to the position of the price of goods. They affect the mechanisms and effects of the exchange of goods in the following way: the tendency to equalize the relative scarcity of factors of production, that is, to achieve uniform prices of the factors of production, is weakened, that is, cannot be realized to the same extent as previously assumed. Since the mobility costs prevent exchange of goods in the cases where these exceed the price difference, they thus also prevent the equalization of prices this exchange of goods would otherwise create. When an equilibrium has been established, there will still be a considerable difference in the prices of production factors in different areas. There is another reason for this

[24] Costs of sales and purchases.
[25] Reduced according to the exchange rate, if the monetary units are different.

difference: mobility costs are now also added to the effects of disproportionate factor supplies still at work even with perfectly mobile goods.

The fact that the tendency to the establishment of uniform prices of goods and factors of production is indirectly and directly obstructed means a loss to the world economy. It is inherent in the meaning of the actual word mobility costs that they involve a sacrifice. Furthermore, the difference in the prices of the factors of production means an uneconomical use of these factors. Both these inconveniences incur a decrease in the profits from the exchange of goods, in comparison to trade at *perfect* mobility of goods. There will still be a profit, however, and it will be larger, the more uniform is the price structure of goods and factors of production, that is, the lower are the mobility costs.

Chapter V

Modifications due to the limited divisibility of factors of production

So far, it has been shown that the difference in the supply of factors of production constitute one reason for exchange between individuals as well as between groups of individuals, so-called regions. This difference, which is only marginally reduced by the mobility of the factors of production or, in other words, the difference in the relative scarcity of factors of production, creates division of labor and exchange.

The question is then whether this is the only reason for exchange. If the factors of production were perfectly mobile, so that each difference in their relative scarcity could be equalized by transfers, would no division of labor then occur? Would each country, each region, each area, each individual procure for himself the goods and services he requires?

The answer must be negative. Division of labor is also due to another factor, that is, increasing profits. Even if the relative scarcity of factors of production were exactly the same in two regions, there would still be division of labor; if each area were to produce all goods it required, this would suppose production on a very small scale.[26] If producing on a small scale, several factors of production would not be fully utilized, however, that is, in those cases where the smallest possible unit is relatively large or relatively less efficient than a larger unit. Since production on a small scale is thus more expensive than production on a larger scale, the latter becomes more profitable. This is usually called the law of increasing returns, and means that there will be one type of production in one area and another type of production in another, and that products will then be exchanged.

This is not only a question of the impact of the law of increasing returns, which means that a firm of a certain size is most suitable financially and, thus, that only a certain number of firms is required; these firms could be spread all over the world and each firm would satisfy its wants in its respective region. An even more

[26] Note the borderline case, where each *individual* produces all his own goods. The production of each good is then small. His numerous occupations prevent him from acquiring skills equaling those of a

considerable division of labor between different regions is due to the fact that two firms or more, which are quite closely located, are often financially superior to those alone in their area, even if their size is the same. Firms which are closely located learn from each other and find it easier to follow technological and organizational progress. Improvements made by one producer stimulate the other in a completely different way if they work on each other's doorsteps than in different parts of the country. They find out about and make use of each other's improvements, for example. A concentration of industries in a certain region thus often leads to an increase in economic efficiency. If so desired, this could be denoted as the limited divisibility of the organizational factors.

Basically, all phenomena of increasing and decreasing returns are due to the limited divisibility of the factors of production.[27] The increasing returns are due to the fact that some indivisible factors of production, of which the firm previously had too rich a supply, are more efficiently used when the scale of production is increased. A tendency to decreasing returns, on the other hand, occurs when an increase in production from one point to another necessitates an acquisition of indivisible factors of production which cannot be completely utilized.

Assume that four factors of production, A, B, C and D, are used to produce a certain good. A is fully utilized at the production rate of 1,000 units of goods a day, B at 500 and C at 200, while D is perfectly divisible, so that the quantity of this good used in production can be freely adjusted to the extent of the production. Furthermore, assume that a unit of A costs 1,000 kronor, a unit of B 500 kronor, a unit of C 200 kronor and a unit of D 1 krona, all costs estimated on a daily basis.

The costs of production at different sizes of firms are then:

---

specialist; this means a certain indivisibility in an individual's productive forces, since specialization, that is, production on a larger scale, makes these more efficient.

[27] The expression decreasing returns is, at times, used for two different phenomena, the one discussed in the text, and the fact that an increase in certain factors of production, for example labor and capital, when the utilization of other factors of production remains constant, leads to a decrease in returns per quantity of labor and capital, cf Bagge. Ek[onomisk] Ts [Tidskrift] 1920.

	Costs					
Production	A	B	C	D	Total	per unit
200 units	1000	500	200	200	1900	8.50
300 “	1000	500	400	300	2200	7.33
400 “	1000	500	400	400	2300	5.75
500 “	1000	500	600	500	2600	5.20
600 “	1000	1000	600	600	3200	5.33
700 “	1000	1000	800	700	3500	5.00
800 “	1000	1000	800	800	3600	4.50
1000 “	1000	1000	1000	1000	4000	4.00
1100 “	2000	1500	1200	1100	5800	5.27
1500 “	2000	1500	1600	1500	6600	4.40
2000 “	2000	2000	2000	2000	8000	4.00
2100 “	3000	2500	2200	2100	9800	4.67

There is thus alternatively an increase and a decrease in the average costs. The optimum is reached at the scale of production allowing a full utilization of all factors of production. A twofold, a threefold increase and so on of this optimum entails no change in the average costs. An increase in production from 500 to 600 units and from

1,000 to 1,100 units incurs increasing costs, that is, decreasing returns, however. In practice, there are no absolute limits to the capacity of one unit of a factor of production. A certain overutilization of a steam-engine, for example, is usually possible, so that production might increase from 1,000 units to 1,050 or even 1,100 units, without the need to acquire either a new unit of A or a new unit of B. Although such an overutilization has certain disadvantages, for example a more extensive use of D (labor) per unit, the average cost might still be as low as for the production of 1,000 units (for example 1,000 (A) + 1,000 (B) + 1,200 (C) + 1,200 (D) = 4,400; 4 kronor per unit). The optimum is thus not a point but rather a range on the scale of production.

Such an overutilization does, however, have its limits. When producing 1,500 units, the production costs are 4.40 per unit, that is, above the optimum. At an increase from 1,000 to 1,500 units, a decrease in returns should be expected. When production increases from 600 to 1,000 units, or from 1,600 to 2,000 units, however, there is a decrease in the average cost.

According to this reasoning, there is an indefinite number of optima, that is, at the production of 1,000*t* units, where *t* is a completely random integer. In practice, however, only a limited number of certain factors of production, above all those related to the management of the firm, can be used in the same firm. There can, for example, be no more than one managing director, and this constitutes a barrier to giant factories.[28] The efficiency of one individual manager decreases when production exceeds a certain limit. Even if the "margin of indeterminateness" is unusually large in this case, that is, a great deal of "overutilization" is possible, productivity will still decrease. The utilization of certain factors of production is thus less efficient. An optimum occurs at the point, or rather in the range, where the average best utilization of the factors of production is obtained. The tendency to a more efficient use when production increases is here met and balanced by the tendency to a less efficient use of certain factors.

[28] What is studied here is the extent of the production in a single firm, technically speaking, not a trust, a cartel or something of the kind.

A special case of limited divisibility occurs when a smaller unit of a certain factor of production can be obtained, but is relatively less efficient than a larger such unit. A steam engine of 1,000 hp does not consume 10 times more coal than a machine of 100 hp. If 1,000 hp had been the smallest possible unit from a technical viewpoint, there would have been pure limited divisibility. This limit is now somewhat obscured since a smaller unit, although less efficient, is technically possible. The disadvantage of only producing 100 hp is thus smaller than if one had to keep a machine with a capacity of 1,000 hp; but it still exists. In this case, we are also dealing with limited divisibility. It is of a different kind, but with essentially the same effects as in the case just dealt with.

It is clear, however, that such a simplified reasoning as above does not completely correspond to reality. Within most industries, the number of firms is much larger than would be required at the most favorable size of firms. Since an extension of production in all firms operating below the optimum would result in decreasing average costs, one might ask why they continue at a smaller scale of production instead of reducing the price and increasing production. It must then be observed that such a process would not be quick and painless and assumes a considerable effort to conquer new markets.[29] Moreover, all firms have equal opportunities. If one firm started selling at its marginal cost, the others would be obliged to do the same and there would be a devastating competition for an unforseeable period of time. For this reason, producers generally prefer not to start any "uncontrolled" competition, for example by selling at their marginal cost on certain markets or temporarily reducing all prices to this level in order to reach an optimum, but will instead maintain a relatively uniform price, covering average costs. The firms which have reached their optimum and thus produce at lower average costs have less reason than others to expand production under these circumstances; they will be satisfied with obtaining a somewhat better return, that is, trade profit.

There seems to be a growing tendency toward such a policy. The aim of the firms to reach an optimum still remains, however. The above-mentioned and other similar tendencies act as checks to this development and are not forces turning it in

[29] cf. Marshall "Industry and Trade", according to which "marketing costs" are of increasing importance.

another direction. Due to the swift changes in all areas of the economy, caused by population growth, technical progress and changes in wants, for example, there will never be a situation resembling an equilibrium, where all firms have reached their optimum.

What are then the consequences of this concentration of production, induced by increasing returns, even if it is not as extensive as might be assumed a priori? In other words, what are the effects of this division of labor and this exchange between different areas or regions? Apparently, there are no other effects than that the inconveniences caused by the limited divisibility of the factors of production are diminished or eliminated. When producing on a small scale, a large number of factors of production were not fully utilized. Due to the concentration of production, production on a large scale and exchange of goods, these indivisible factors of production are also fully or at least almost fully utilized.[30]

The closer one is to an optimum, the more complete is utilization. The increase in exchange thus means that production is concentrated in a small number of areas, each with a relatively large-scale production. The factors of production are thus better utilized and there are fewer disadvantages due to limited *divisibility*. This tendency of exchange to remove the disadvantages of the limited *divisibility* of the factors of production exactly equals the tendency to remove the disadvantages resulting from their *limited* mobility dealt with earlier, since it entails a uniform price structure of these factors for their most efficient use. There is thus a tendency for exchange to create a situation where the utilization of the factors of production remains the same, or rather, as efficient, as if perfect divisibility and mobility did exist. The gain is then that the most inefficient use of the factors of production is prevented or reduced.

Just as exchange cannot create a completely uniform price structure, due to a certain disproportion in the distribution of the factors of production and their utilization in production, that is, just as the effects of limited mobility cannot be

[30] It is the *static* price formation theory and the effects of the limited divisibility on this theory that are studied here, although it has been considered that a dynamic factor should be included in the reasoning. No account is taken of such dynamic phenomena as regular or irregular variations in demand, however, which prevent a full utilization of indivisible factors of production.

entirely removed, another phenomenon of disproportion will counteract the tendency to eliminate the effects of limited divisibility.

The optimum size of a firm is thus not the point where all factors of production are fully utilized, but the point where the overall utilization is the best possible. If all factors were fully utilized, the scale of production would often be so large that it would be uneconomical for other reasons.[31] At the optimum, the advantages of extending production and thus, obtaining a better utilization of certain factors of production will be counterbalanced by the disadvantages of mass production, such as less efficient management, more expensive control and the employees having less power of initiative.

Even if all firms were at an optimum, this would not necessarily entail a full utilization of the factors of production but would only mean that the disadvantages of limited *divisibility* were minimized. As already mentioned, if an on-the-spot picture of the economy was presented, it would show that a number of firms fall below the optimum. One of the reasons for this has already been mentioned. Whatever these reasons might be, however, the actual existence of such firms means that division of labor and exchange has not been carried so far that optimum utilization occurs. Besides the phenomenon of disproportion just mentioned, this is another circumstance which counteracts the tendency of exchange to remove the effects of limited divisibility. There is, however, also another obstacle.

In order not to complicate the reasoning more than necessary, the mobility costs of the goods have not been accounted for so far in this chapter, although they clearly constitute an obstacle to the division of labor and exchange. The optimum size of a firm does not only depend on the pure costs of production, but exists at the point where an increase in production incurs new costs, which together with the transport costs exactly cover the highest price to be obtained on a possible new market. Assume, for example, that 1 million units are produced at the average cost of 4 kronor and that an increase in production of 100,000 units would incur costs of 350,000 kronor. Returns are then increasing, that is, the average costs are falling. Nevertheless, this expansion of production will incur a loss, if the highest price the

[31] Cf the above concerning obstacles to giant firms.

goods can command on the new market, without inviting competition, is 4.50 kronor and transport costs are 1.15 kronor per unit. The cost price on this market is thus 3.50 + 1.25 = 4.75. The optimum is thus a production of 1 million units, despite the fact that the average production costs would fall if production were expanded. This real optimum apparently exists at the point where the advantage of decreasing production costs are equalized by the disadvantage of increasing transport costs, or rather, at the point where the prices on possible new markets less transport costs begin to fall below the additional costs for an increase in production.

If sales of a product from one region to another are to take place, the increase in costs in the export region together with transport costs must thus fall below the production costs in the other region. This condition might very well be fulfilled, even if the relative prices of the factors of production are equal. Assume that A can produce good *a* as well as good *b* at the price of 1 krona per unit; and so can B. Furthermore, assume that if A also produces the good for B, the additional costs will only amount to 0.75 kronor per unit and the good can thus be sold to B at that price. In the same way, B can sell good *b* to A at the price of 0.75 kronor. As long as transport costs fall below 0.25 kronor, it is not profitable for A to produce good *b* or for B to produce good *a*; instead division of labor and exchange will take place.

The *precondition* for exchange is thus that a more efficient use of the non-divisible factors of production more than outweighs the disadvantages of transporting the goods. The *function* of the exchange is thus to obtain a more efficient use of the factors of production, to which transport costs constitute an obstacle.

It has already been noted that division of labor and exchange are not as extensive as would have been required by the prospects of falling costs due to expanded production; that is, many firms operate much below the optimum. This will now be somewhat further illustrated in connection with the mobility costs of the goods.

Reaching the optimum production of a good, for which the transport costs vary considerably with the distance,[32] means, according to the above result, that the producers apply price discrimination. When deciding whether a certain sale to a new market is to take place, only the marginal costs must be considered, that is, the increase in total costs which a required increase in production would incur. For various reasons – one of which has already been mentioned[33] – many firms apply uniform prices,[34] approximately corresponding to the average costs of an average firm.[35] The best firms, that is, those which are close to the optimum, are thus profitable, while the least successful incur a loss; assuming "normal" business cycles.

Failing to use price discrimination means that the concentration of production and exchange is not taken as far as required by the above mentioned conditions.[36] Even if a certain quantity of a certain type of good can be produced at the additional cost of, for example, 3 kronor per unit, the transport costs to a "new" market are 1 krona and the price on this market 4.50 kronor per unit, sales to that market will generally not take place, if the average costs of the firm are 4 kronor and can only be reduced to 3.90 kronor per unit through the expansion. Failing to use price discrimination thus works as a check on the firm's tendency towards the optimum.

A summary of the results of Chapter 5 is that increasing returns and the resulting exchange of goods, are due to the limited divisibility of the factors of production. They might also occur even if the relative prices of the factors of production in two regions are the same. A full utilization of the non-divisible factors of production is made possible by the exchange of goods. Full utilization seldom occurs, however.

1:0/the mobility costs of the goods prevent all kinds of exchange, unless the profits from a more efficient use of the factors of production are higher than the mobility costs.

---

[32] For a number of goods, it is of little or no importance if they are transported 20 or 1000 kilometers. Packing and shipment costs are in no way related to distance. An increase in other costs might be insignificant, expressed as a percentage.
[33] Another reason is that implementing price discrimination is, in practice, very difficult.
[34] There are exceptions, of course, for example within power producing firms. Price discrimination in sales to other countries will be studied in the next chapter.
[35] cf Marshall's "representative firm".
[36] See above.

2:0/ not even if the mobility costs of the goods were removed would *all* factors of production be fully utilized in a firm; a certain disproportion in their composition cannot be avoided, as long as the firms cannot be expanded to any extent.

3:0/ Not all firms reach their optimum size, on account of the two factors stated above.[37]

[37] This dynamic factor is included already at this stage, so that the argument does not rest on excessively unrealistic conditions.

Chapter VI

The theory of interregional exchange without simplified assumptions

Summary

It is now time to address the issue of the conditions and effects of interregional exchange, on basis of the previous observations, without any simplifying assumptions; that is, to try to present a picture of the price formation mechanism, the way it works with limited mobility and divisibility of the factors of production and limited mobility of the goods.

The studies have shown that there might be two reasons for exchange; the relative scarcity of factors of production, on the one hand, and increasing returns, on the other which, in turn, are connected to the limited divisibility of the factors of production. The exchange which then takes place tends to equalize the relative scarcity of factors of production and make a more complete utilization of the indivisible factors of production possible, in other words, create a more economic utilization of the factors of production. The tendency is to obtain as efficient a use as if these had been perfectly mobile and divisible. To the same extent that this tendency is realized, the difference in the relative scarcity of factors of production will decrease and disappear, and the degree to which they are utilized will increase and become the largest possible, since the industries will approach and attain the optimum. The exchange thus removes its own causes. Realizing this tendency makes all *new* exchange impossible, while the ongoing exchange continues.

The tendency cannot be fully realized, however, partly due to certain phenomena of disproportion, and partly to obstacles to the mobility of goods, mainly transport costs. In addition, the equilibrium which would appear through the hindrance of this tendency by these two obstacles, does not exist in practice. If one switches from a purely static approach and also considers dynamic factors, it is found that the attempts of the economy to reach this equilibrium will never be fully realized, due to continuous changes in the determining factors. Division of labor and exchange

do not attain either the direction or the extent which would give the highest economic efficiency.

The factors studied here – even if disregarding the dynamic factor – depend on a modification of the price formation theory, building on the assumption of perfect mobility and divisibility. In this theory, the total supply of factors of production, the technical conditions[38] and the hierarchy of wants of the population can be considered as the only factors determining price formation. Price formation in the isolated economic unit that the world constitutes does not have such a simple structure.

The different supply of factors of production in different areas, their limited mobility and divisibility and the limited mobility of the goods[39] are factors determining a specific location of production. They are thus part of the determinants of the price formation system. Due to their existence, the prices of goods as well as of factors of production are completely different than would otherwise have been the case. As shown from this study, exchange does mean that there is a tendency to create the same price structure as if perfect mobility and divisibility did exist. For various reasons, this tendency cannot be fully realized. These factors will thus have a considerable effect on the price structure.

Besides the total quantity of the factors of production, their local distribution is of importance. The technical conditions must be understood in another and more comprehensive context, which also includes the limited divisibility of the factors of production and the suitability of land for transportation. For each possible state of these factors and the hierarchy of wants of the population, there is a specific production pattern: the location of production as well as the prices of the factors of production and the goods are determined, that is, the entire price formation system is determined.

[38] See appendix 3 [appendix is missing].
[39] That is, the properties of commodities and geography with respect to [the facility of] transport.

SEMINAR PAPER SERIES

The Series was initiated in 1971. For a complete list of Seminar Papers, please contact the Institute.

1996

616.	*Assar Lindbeck:*	The West European Employment Problem. 31 pp.
617.	*Assar Lindbeck:*	Full Employment and the Welfare State. 22 pp.
618.	*Javier Ortega:*	How (Good) Immigration Is: A Matching Analysis. 30 pp.
619.	*Joakim Persson and Bo Malmberg:*	Human Capital, Demographics and Growth Across the US States 1920-1990. 21 pp.
620.	*Assar Lindbeck and Dennis J. Snower:*	Centralized Bargaining, Multi-Tasking, and Work Incentives. 43 pp.
621.	*Paul Söderlind and Lars E.O. Svensson:*	New Techniques to Extract Market Expectations from Financial Instruments. 47 pp

1997

622.	*Assar Lindbeck:*	Incentives and Social Norms in Household Behavior. 12 pp.
623.	*John Hassler and José Vicente Rodriguez Mora:*	Employment Turnover and Unemployment Insurance. 36 pp.
624.	*Nils-Petter Lagerlöf:*	Strategic Saving and Non-Negative Gifts. 20 pp.
625.	*Lars E.O. Svensson:*	Inflation Targeting: Some Extensions. 43 pp.
626.	*James E. Anderson:*	Revenue Neutral Trade Reform with Many Households, Quotas and Tariffs. 36 pp.
627.	*Mårten Blix:*	Rational Expectations in a VAR with Markov Switching. 37 pp.
628.	*Assar Lindbeck and Dennis J. Snower:*	The Division of Labor Within Firms. 12 pp.
629.	*Etienne Wasmer:*	Can Labour Supply Explain the Rise in Unemployment and Inter-Group Wage Inequality in the OECD? 64 pp.

630. *Torsten Persson and Guido Tabellini:* Political Economics and Macroeconomic Policy.100 pp.

631. *John Hassler and Assar Lindbeck:* Intergenerational Risk Sharing, Stability and Optimality of Alternative Pension Systems. 38 pp.

632. *Michael Woodford:* Doing Without Money: Controlling Inflation in a Post-Monetary World. 62 pp.

633. *Torsten Persson, Gérard Roland and Guido Tabellini:* Comparative Politics and Public Finance. 55 pp.

634. *Johan Stennek:* Coordination in Oligopoly. 14 pp.

1998

635. *John Hassler and José V. Rodríguez Mora:* IQ, Social Mobility and Growth. 34 pp.

636. *Jon Faust and Lars E. O. Svensson:* Transparency and Credibility: Monetary Policy with Unobservable Goals. 40 pp.

637. *Glenn D. Rudebusch and Lars E. O. Svensson:* Policy Rules for Inflation Targeting. 51 pp.

638. *Lars E. O. Svensson:* Open-Economy Inflation Targeting. 51 pp.

639. *Lars Calmfors:* Unemployment, Labour-Market Reform and Monetary Union. 35 pp

640. *Assar Lindbeck*: Swedish Lessons for Post-Socialist Countries. 37 pp.

641. *Donald Brash:* Inflation Targeting in New Zealand: Experience and Practice. 11 pp.

642. *Claes Berg and Lars Jonung:* Pioneering Price Level Targeting: The Swedish Experience 1931-1937. 50 pp.

643. *Jürgen von Hagen:* Money Growth Targeting. 34 pp.

644. *Bennett T. McCallum and Edward Nelson:* Nominal Income Targeting in an Open-Economy Optimizing Model. 48 pp.

645. *Assar Lindbeck:* Swedish Lessons for Post-Socialist Countries. 42 pp.

646. *Lars E.O. Svensson:* Inflation Targeting as a Monetary Policy Rule. 51 pp.

647. *Jonas Agell and Mats Persson:* Tax Arbitrage and Labor Supply. 35 pp.

648. *Frederic S. Mishkin:* International Experiences With Different Monetary Policy Regimes. 47 pp.

649. *John B. Taylor:* The Robustness and Efficiency of Monetary Policy Rules as Guidelines for Interest Rate Setting by The European Central Bank. 39 pp.

650. *Christopher J. Erceg, Dale W. Henderson and Andrew T. Levin:* Tradeoffs Between Inflation and Output-Gap Variances in an Optimizing-Agent Model. 43 pp.

651. *Etienne Wasmer:* Labor Supply Dynamics, Unemployment and Human Capital Investments. 36 pp.

652. *Daron Acemoglu and Fabrizio Zilibotti:* Information Accumulation in Development. 43 pp.

653. *Argia Sbordone:* Prices and Unit Labor Costs: A New Test of Price Stickiness. 33 pp.

654. *Martin Flodén and Jesper Lindé:* Idiosyncratic Risk in the U.S. and Sweden: Is there a Role for Government Insurance? 30 pp.

655. *Thomas P. Tangerås:* On the Role of Public Opinion Polls in Political Competition. 36 pp.

656. *Peter Svedberg:* 841 Million Undernourished? On the Tyranny of Deriving a Number. 39 pp.

657. *Lars Calmfors:* Macroeconomic Policy, Wage Setting and Employment – What Difference Does the EMU Make? 52 pp.

658. *Torsten Persson and Guido Tabellini:* The Size and Scope of Government: Comparative Politics with Rational Politicians. 47 pp.

659. *Lars Calmfors:* Monetary Union and Precautionary Labour-Market Reform. 10 pp.

660. *Daron Acemoglu and Fabrizio Zilibotti:* Productivity Differences. 48 pp.

661. *Ramon Marimon and Fabrizio Zilibotti:* Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits. 35 pp.

662. *Yves Zenou:* Urban Unemployment and City Formation. Theory and Policy Implications. 35 pp.

663. *Stefan Palmqvist:* Why Central Banks Announce Their Objectives: Monetary Policy with Discretionary Signalling. 22 pp.

664. *Kjetil Storesletten:* Sustaining Fiscal Policy Through Immigration. 41 pp.

665: *John Hassler, José V. Rodríguez Mora, Kjetil Storesletten and Fabrizio Zilibotti:* Equilibrium Unemployment Insurance. 56 pp.

1999

666. *Michael Woodford:* Optimal Monetary Policy Inertia. 112 pp.

667. *Lars E.O. Svensson:* Monetary Policy Issues for the Eurosystem. 54 pp.

668. *Assar Lindbeck:* The Price in Economic Sciences In Memory of Alfred Nobel 1969-1998. 28 pp.

669. *Lars E.O. Svensson:* The Equilibrium Degree of Transparency and Control in Monetary Policy. 25 pp.

670. *Assar Lindbeck and Solveig Wikström:* The ICT Revolution in Consumer Product Markets. 23 pp.

671. *Lars E.O. Svensson:* Does the *P** Model Provide Any Rationale for Monetary Targeting? 14 pp.

672. *Glenn Rudebusch and Lars E.O. Svensson:* Eurosystem Monetary Targeting: Lessons from U.S. Data. 29 pp.

673. *Lars Svensson:* Price Stability as a Target for Monetary Policy Defining and Maintaining Price Stability. 50 pp.

674. *Thomas P. Tangerås:* Collusion-Proof Yardstick Competition. 33 pp.

675. *Bertil Ohlin:* The Theory of Interregional Exchange. 42 pp.

ISSN 0347-8769
Stockholm, 1999
Institute for International Economic Studies