LOVELY
PROFESSIONAL
UNIVERSITY

THEORY OF INTERNATIONAL TRADE

Edited By

Dr. Pavitar Parkash Singh

SYLLABUS

Theory of International Trade

Objectives:

The course provides an understanding of international trade and investment theories. It is designed to better understand the implications of such theories as they relate to international trade management. It helps students deal with the opportunities and challenges created by the global environment.

Sr. No.	Content
1	Trade as an engine of growth, Measurement of gains from trade, Free Trade Theory-Absolute advantage, comparative advantage & opportunity cost, Modern theories of international trade: Theorem of factor price equalization, H-O Theory, Kravis & Linder theory of trade.
2	Role of dynamic factors : tastes, technology & factor endowments in trade, Rybszynski Theorem, Causes of emergence & measurement of intra industry trade and its impact on developing economies,
3	Tariff, Quotas & non-tariff barriers: Definitions & types, Economic effects of Tariff & Quotas on national income, output & employment, Political economy of non-tariff barriers and their implication.
4	Balance of Payments and Balance of Trade: Meaning & components, Equilibrium & Dis-equilibrium in BOP, BOP Adjustment: Monetary approach, Exchange Rate: meaning & components,
5	Theories of Determination of Exchange rate (PPP, Monetary), Theories of Determination of Exchange rate (Portfolio & Balance of Payment), Process of adjustments : Gold standard, Fixed Exchange Rates & Flexible Exchange Rates, Merits & demerits of Fixed & Flexible exchange rate.

CONTENTS

Unit 1 : Trade as an Engine of Growth

```
CONTENTS
Objectives
Introduction
  1.1  A Brief Historical Sketch
  1.2  Classical Period : International Trade and Growth
  1.3  Post Classical Period : International Trade and Growth
  1.4  Models of Endogenous Growth and International Trade
  1.5  Summary
  1.6  Key-Words
  1.7  Review Questions
  1.8  Further Readings
```

Objectives

After reading this Unit students will be able to:

*   Describe the Classical and Post Classical Period of International Growth and Trade.

*   Explain the Models of Endogenous Growth and International Trade.

Introduction

The promotion of free trade is one of the oldest policy implications offered by international economic theory. While significant disconnects have historically existed between the politics and the economics of trade policy, the rapid economic growth experienced by the export-oriented Asian countries during the 1960s and 1970s amidst a largely stagnating and trade-restrictive developing world provided a precedent for effective development policy, especially within the world's less developed countries (LDCs). Free trade arguments have since been championed by a majority of global institutions, including the International Monetary Fund (IMF hereafter), the International Bank for Reconstruction and Development (IBRD) or World Bank, the Organization for Economic Co-operation and Development (OECD hereafter), and the World Trade Organization (WTO hereafter). The argument goes that a reduction in trade barriers will induce greater economic efficiency within LDCs by offering cheaper world prices to domestic consumers (increasing consumer welfare) while creating conditions of competition for domestic producers (forcing domestic production to shift towards the most efficient sectors based upon availability of domestic factors).

1.1 A Brief Historical Sketch

It can be said that the positive effects of International Trade (IT) on Economic Growth (EG) were first pointed out by Smith (1776). This idea prevailed until World War II (WWII), although with relative hibernation during the 'marginalist revolution'. After WWII, the introverted and protectionist EG experiments had some significance, especially in Latin America. From the 60's on, owing to the failure of those experiments and to the association of quick EG with the opening of IT and the consequent international specialization in several countries, as well as to the results of many studies based on the neoclassical theories of EG and IT, a new decisive role was given to IT as EG's driving force.

However, although the dominant theoretical position tended, from the beginning (with the Classics), to indicate a positive relation between IT and EG, many studies linked the gains of IT only with static effects. But Baldwin (1984), for example, concluded, in a survey of empirical studies, that the static effects were of little significance. The debate has widened in the last decades, precisely in the direction

of pointing out and stressing the dynamic effects of IT. The theoretical development afforded by the models of endogenous EG [especially after the works of Romer (1986) and Lucas (1988)], which stimulated the creation of empirical studies, moved toward an integrated analysis of the EG and IT theories. So, the classical tradition, apparently interrupted by the neoclassical separation of those two areas of the theory, seems to have been recovered, assigning, as a result, a decisive role to IT on the countries' rate of EG.

The recognition of this importance has even led to the ceaseless appearance of proposals from international organisations, such as the World Bank (WB) and the United Nations (UN). As a result, many countries began to reduce commercial barriers and other controls of economic activity and obtained a significant (and lasting) increase in the rate of EG, which suggests that extroversion has a dynamic effect on the economy, helping to speed up the rate of EG. Moreover, the processes of economic integration intensified.

Aims and Structuring of the Work

The EG theory analyzes, at an aggregate level, the evolution of the real product and its distribution (intra and inter countries). In general, the models regard that product as created with a limited and aggregate number of factors. Models which are initially designed to explain the EG of the Developed Countries (DCs) are, in general, 'supply side' models because it's admitted that, in the long-term, the product of equilibrium is located in the proximity of the potential product, and because the latter depends on the availability of the factors and technological level. The main objectives of those models are to explain the variations of the factors and of the production function itself (*i.e.*, of the way on which the product depends on the factors) and account the effects that these variations have on the evolution and distribution of production.

Our aim is to analyze the impact of commercial and technological effects (ignoring the financial component), resulting from IT, on the physical accumulation of productive factors and on its improvement (efficiency gains). In other words, in the rate of EG, during the evolution of economic growth theory. We then underscore studies that manifestly convey the 'effect of EG' (changes that modify, in a durable way, the rates of EG and its tendency in the long-term), instead of simple 'level effects' (changes that influence the EG only in the short-term).

Did u know? Ricardo (1817) presented a 'dynamic model of EG' with three forces and two restrictions. He characterized the progressive states as having high savings, capital accumulation, production, productivity, benefits and labour demand forcing the increase of wages and demographic growth.

The structure which is followed in this paper observes the temporal evolution and the status that we think commercial and technological aspects have in what concerns the EG models. In effect, it seems to us that in the 'classical period' the EG and IT theories were linked (section 2), that in the 'neoclassical period' there was a tendency toward their separation (section 3), and that recently, with the new endogenous EG approaches, they were again considered jointly (section 4). Finally, in section 5 we present the main conclusions.

1.2 Classical Period : International Trade and Growth

Since the classics don't distinguish the questions of EG from the questions of IT, the examination of this problem leads us to the classics' main models of IT. However, given the aim of this work, we attempt to advance on those models which basically discuss the 'static gains of the IT'.

As far as the interaction between IT and EG is concerned, we found two main ideas to point out in Smith (1776). On the one hand, IT made it possible to overcome the reduced dimension of the internal market and, on the other hand, by increasing the extension of the market, the labour division improved and the productivity increased. The IT would therefore constitute a dynamic force capable of

intensifying the ability and skills of workers, of encouraging technical innovations and the accumulation of capital, of making it possible to overcome technical indivisibilities and, generally speaking, of giving participating countries the possibility of enjoying EG.

However, in view of the limitations of land, both in quantity and in quality, the additional alimentary resources were obtained in conditions of decreasing returns, in which the production is absorbed by wages in an increasing proportion, reducing the stimulation of new investments and, sooner or later, reaching the 'stationary state'. IT could delay the fall in the rate of profit. Apart from the contribution of IT, underestimating the importance of technology, he underestimated the positive effects of IT on technology.

Finally, among the Classics, Mill (1848) also explicitly reported the Classic point of view according to which the production resulted from labour, capital, land and their productivities. And just like Ricardo, he recognized that underlying the 'progressive state' there was the 'stationary state', and that ultimately the force capable of delaying this state was technical progress. Accordingly, the emphasis that Smith had placed on the extension of the market decreases, even though he also defended free trade among countries. We think that this situation was the result of the expectation created by the Industrial Revolution (IR) in regards to technical progress.

1.3 Post Classical Period : International Trade and Growth

The structure of this section takes into account the separation that occurred between IT and EG theories, and takes also into consideration some reactions to the classical and neoclassical theories. We begin with the neoclassical IT theory, proceed to the post-classical EG, before Solow, and then go on to the reactions. Afterwards comes the modern neoclassical theory of EG, and we conclude with the disclosure of extensions or works of synthesis, applications, and studies of commercial policies that discuss the theme under analysis.

Neoclassical International Trade

The followers of Ricardo ignored the question of the foundations of comparative advantages and didn't identify factors, resulting from IT, that could raise, in a lasting form, the rate of EG and its tendency in the long-term. In general, the changes introduced in the ricardian theory demonstrated the increase of welfare caused by IT, but ignored eventual gains in the rate of EG. It was in the context of neoclassical general equilibrium that the model of Heckscher (1919) and Ohlin (1933) appeared, whose contributions Samuelson (1948 and 1949) completed in the late 40's. In a rigid analysis of the model, we observe that it permits to advocate the opening of the countries to IT, showing that it is efficient, mutually beneficial and positive for the entire world. However, it limits the analysis to the static gains of welfare.

Post-classical Growth, before Solow

Generically, the classical economists gave us an idea of the race between the increase of the population and EG, with an uncertain winner. This version gradually disappeared with the IR, because the product increased from decade to decade in increasingly larger areas. That might be the reason why EG was no longer seen as a problem and why it wasn't amply pursued in the studies and writings of the following economists.

Notes Classical thought gave way to 'marginalism' from the 1870s onwards. This fact led to a 'new theory' (neoclassical) which, for some time, kept the main lines of the evolution of the economy in the long-term away from the studies.

Nevertheless, Marshall (1890) pointed out that "The causes which determine the economic progress of nations belong to the study of international trade". In effect, the expansion of the market that it

represented led to the increase of global production and originated the increase of internal and external economies, which resulted in increasing income for the economy. But, although he understood the importance of those externalities, he also recognized the difficulties of his analytic treatment. Among his successors, only Young (1928) was concerned with EG when he considered, like Smith, that the dimension of the market limited the labour division (and therefore, the productivity). He also examined the inter-relation between industries in the process of EG, the creation of new industries due to the specialization resulting from the extension of the market, the importance of specialization and standardization in a vast market and the influence of this market on technological progress.

Another exception of this period's remarkable was Schumpeter (1912, 1942 and 1954), who repeated old points of view concerning the tendency of the profit to reach a minimum and the dependency of the rate of EG on capital accumulation. But he went further, distinguishing 'invention' (advancement of useful knowledge to production) from 'innovation' (economic activity of exploring that knowledge). Considering the latter as the central element of EG, he described the exigencies for a successful innovation, which included the need for markets opened to the exterior.

We conclude this subsection by mentioning some authors who made the restart of studies of dynamic themes — and, consequently, of the EG theory — easier, thus laying a good foundation for future investigations. Ramsey (1928) introduced the description of EG and the principle of research of an optimum EG. Cobb and Douglas (1928) presented production functions that became known as *Cobb-Douglas* production functions and which constituted an essential element of numerous models of EG. Harrod (1938 and 1948) and Domar (1937 and 1946) independently developed a model inspired in Keynes, which gave the research of EG an important momentum and a specific direction. Finally, Rosenstein-Rodan (1943) retrieved some of Young's ideas, when the problems of the Less Developed Countries (LDCs) attracted the economists' attention.

Reactions of Classical and Neoclassical Theories

Immediately after the end of WWII, the dominant position was questioned, namely in the case of the LDCs. Those reactions abandoned the classical and neoclassical orientation in considering hypotheses that were strange to them. The introverted and protectionist EG experiments of Latin America (industrialization for import substitution) also stood out, with rationalization and justification owing, first of all, to some structuralist economists [Prebisch (1949) — executive secretary of UN — and Singer (1950)] and to the UN Economic Commission for Latin America (ECLA). Essentially, they defended that the IT brought on negative consequences in the long-term for the LDCs because their specialization occurred in products with low demand income elasticity and, therefore, with a weak perspective of exports growth, and noticed a tendency for the constant deterioration of trade terms. Furthermore, this specialization entailed significant economic and social costs of adaptation to the evolution of the chain of IT.

Myrdal (1956 and 1957) sustained that IT didn't equal the remuneration of factors (in contradiction with the proposal of the neoclassical model) and that, unlike the industries of the DCs, the traditional industries of the LDCs remained weak. In short, the IT had some positive effects of diffusion on the LDCs, but in the long-term the negative effects remained because it stimulated a production of primary goods (plantations and mining enclaves) subject to irregular prices and demand. Lewis (1954 and 1969) and the marxist author Emmanuel (1969) decided, respectively, on the deterioration of the trade terms of the LDCs and on the existence of unequal trade biased against the LDCs. Nurkse (1959) also questioned the relevance of commercial trade between the DCs and the LDCs for the latter. Perroux (1978) considered that the LDCs were controlled. Consequently, the EG and the structural transformation were induced by the DCs, which will cause the loss of potential positive effects to the external world, in the long term.

Another group of (radical) authors observed the economic relations as a whole (chain of goods, services and capitals) : radical marxist visions [among others, Destanne de Bernis (1977) and Andreff (1981)] and the dependency theory [among others, Santos (1970), Frank (1970) and Amin (1970 and 1973)]. Basically, they defended that the underdevelopment was the consequence of the changes and deformations in the economic and social structures caused by the economic and social relation that existed with DCs.

Modern Neoclassical Theory of Growth

In the late 50's and early 60's the interest for the EG reawakened with the recovery of the classical approach, according to which the production was a function of labour, capital, land and their productivities. The question of the 'accounting of EG' was also raised.

We can be pointed out 1956 as the year of birth of the 'modern neoclassical theory of EG' with Solow [and Swan (1956)]. The proposed model describes the relation between savings, accumulation of capital and EG based on a function of aggregate production (crucial supply), and there was a point of sustainable equilibrium (steady-state), which would be reached regardless of initial conditions. By increasing the productivity of the factors, the exogenous technical progress created positive effects on the process of accumulation and made the model compatible with a balanced growth path. In economic terms, this means that it took into account the convergence between economies. Moreover, along with the diffusion of technical progress there would be a convergence of the rate of EG per capita for a common steady-state. Consequently, it can be said that, by facilitating the diffusion of technical progress, the IT would be important for the LDCs.

As far as the 'accounting of EG' is concerned, Solow (1957) used the function of aggregate production as a starting point to measure the sources of EG in the United States. The rate of EG springs from labour and capital growth rates (which we call traditional sources), weighed by the respective participation in production and technical progress or total productivity of factors (TPF). The TPF resulted from the difference between the observed rate of EG and the part of that EG explained by the traditional sources (thus the designation 'residual of Solow'). Clearly he distinguished 'EG effects' (the three sources mentioned above) from 'level effects'. As a result, IT would, eventually, be a 'level effect' that would create positive effects in a transitory period of time.

Task   What do you mean by Modern Classical Theory?

From Solow on, many economists considered the advance of knowledge to be a source of the 'residual'. However, the 'accountants of EG' (post Solow) included as sources the contributions of many elements such as the accumulation of 'human capital', economies of scale, the improved allocation of resources and the new generations of more productive machines [among others, Kendrick (1961), Denison (1962, 1974 and 1985) and Griliches and Jorgenson (1967)]. However, they didn't quantify the advancement in knowledge, leaving a residual factor unexplained. Furthermore, they didn't include IT, at least not explicitly, as a source of EG. We think that this situation is due to two factors that have already been mentioned. On the one hand, the separation that occurred between the theories of IT and EG, and on the other, the effects of IT on the level and not on the long-term rate of EG.

Theoretical Synthesis, Empirical Applications and Commercial Policies

As we have said, the works of the 'accounting of EG' widened the scope of studies of the sources and began studying different structural situations, abandoning therefore some neoclassical assumptions. Thus, studies done since the late 1960s considered, besides the traditional factors, other explanatory variables, maintaining the functional scheme proposed by Solow. In this context, in view of the need to determine the totality of growth sources and in view of the failure of introverted growth experiments, along with EG's association with the opening of IT, there was an increase in the research on trade and growth.

We present some theoretical studies and empirical applications which ensued, as well as studies/ recommendations on the external commercial policy, whose defining characteristic resides in the fact that IT (above all the exporting component) is considered an explanatory variable of EG. They generally associate this situation with an improved allocation of resources (according to the comparative advantages), with a greater utilisation of productive capacity (which makes it possible to obtain economies of scale), with a greater propensity to implement technological improvement (in answering to the greater competition that they are subjected to), and with the higher level of employment created when compared to introverted strategies.

Theoretical synthesis

We begin with the structuralist synthesis of EG of Kuznets (1972), Chenery and Syrquin (1975 and 1989) and Chenery *et al.* (1986). In brief, we noticed that what is most relevant is the fact that the observation of the process of EG of the country depends on the changes of factorial provision but also, and especially, on changes in demand, leading to the increase of the internal market, the substitution of imports and the variations of exports. In this sense, they defend that the TPF included, among other factors, the ones associated with the weight and countenance of IT.

In turn, in a brief reference to the analyses that underscore economic integration (more or less institutional), we mention, for instance, Young (1928), Florence (1948), Stigler (1951), Meade (1953), Svennilson (1954) and Scitovsky (1958). This group of authors took dynamic effects into account, namely those resulting from the increase in competition, from the gain of economies of scale, from changes in the level and nature of investments, from the increase of research expenses, from technical progress and from the elimination of the risk and uncertainty in trade.

Another example is Findlay's model (1980 and 1984) for the commercial relations between the (developed) North and the (underdeveloped) South. While integrating the neoclassical theories of IT and EG and at the same time recognizing the specificities of the LDCs, he assumes that the economy of the North is dynamically described by Solow's (1956) model of EG, except for the fact that it consumes an importable good in addition to its own product, while the economy of the South works according to Lewis' (1954) model of unlimited supply of labour. The terms of trade [based on Johnson (1967)] related EG in the two economies. So, the South had the IT as the principal driving force of EG. However, the rhythm of EG was determined by the (exogenous) EG rate of the North.

We conclude with the work of Feder (1982), where EG proceeded from the effects of the traditional sources and from the exporter sector performance. In brief, he considers that economies have two distinct productive sectors (exporter and non-exporter), differing in the final destination of productions and in the superiority of the productivity of the traditional factors in the exporter sector. He concluded that the rate of EG was explained by the rates of investment, labour growth and exports growth. He also presents a way of comparing the relative benefits of the allocation of resources to both sectors.

Empirical applications

In what concerns empirical applications, we immediately point out the structuralist inclination present in Hagen and Hawrylyshyn (1969), Chenery *et al.* (1970), Chenery *et al.* (1986) and Chenery and Syrquin (1989). These authors tested the significance of 'structuralist' variables, and decided on its relevance in explaining EG, particularly in samples of LDCs and in the years that followed the 60's. They demonstrate, with empirical studies, the evidence that the exports promote EG. Moreover, they claim that the existence of imports limits may reduce EG.

Feder (1982) proceeded with the empirical application of the developed framework, in semi-industrialized and marginally semi-industrialized countries, between 1964-1973.

He concludes that, statistically, its formulation was superior to the traditional neoclassical formulation. He also decided on the superiority of the marginal production of the factors in the exporter sector and on the externality of this sector over the other. Finally, he concluded that the allocation of one unit of capital to the exporter sector would create one marginal value for the economy superior to what would be obtained if it were affected by a non-exporter sector. Ram (1987) extended the analysis of Feder to the estimation of time-series for each country from a sample of 88 LDCs, in the years 1960-1985. The obtained regressions (being globally statistically significant) confirm the positive effect of the exporter sector, in about 70% of the countries.

We conclude by saying that even more sceptic empirical applications like those of Michaely (1977), Tyler (1981) and Dodaro (1991) do not challenge the positive effect of IT on EG, provided the countries have reached a certain minimum threshold of development.

The question of the international trade policies

In view of the failure of introverted EG experiments, of the success of extroverted EG experiences (case of countries of Southeast Asia) and of the dominant theoretical thought, the UN started to

recommend the opening to IT. They started the process with resolution 1707 of 1961 and continued, for example, in 1964 with the UN conference on trade and development (UNCTAD I). Both the General Agreement on Tariffs and Trade (GATT/WTO), through successive rounds of negotiations and the recommendation of the Organisation for Economic Cooperation and Development (OECD) worked also in favour of the liberalization of trade [see, for example, Arndt (1987, pp. 72-77)].

Little *et al.* (1970) considered the strategy of substitution of imports to be responsible for the existence of firms with high costs, charging consequently high prices for their products, which can only be purchased by high income consumers. Thereupon this situation would lead to the dependence of the enterprises on governmental decisions. Therefore they defended the promotion of exports.

Balassa (1978) compared the strategies of promotion of exports with those of substitution of imports. His work is based on Michalopoullos and Jay (1973). He considers a sample of 10 LDCs with different grades of use of those strategies (in 1960-1966 and 1966-1973). Taking neoclassical production function, he uses different versions functional forms, resulting from different exporting performances. From the results, he stressed, on the one hand, the significance of the export growth and, on the other hand, that the countries with rates of export growth higher than the average also registered the best performances. More recently, Balassa (1986 and 1987) analyzed the EG, between 1963-1984, of a group of LDCs that he divided in those turned toward the exterior and turned toward the interior, concluding that the former exceeded the performance of the latter, especially from the middle of the 70's on.

In 1985, Krueger observed that especially from the early 60's on, some LDCs reduced commercial barriers and other controls of economic activity and obtained a significant (and lasting) increase in the rate of EG. Namely, technological factors, of economic behaviour and political and economic consideration that involved dynamic effects (besides the static effects), helped explain the differences of performance among economies. Rajapatirana (1987), co-responsible for the *World Development Report* 1987, claimed again Krueger' arguments, considering that the IT allowed for dynamic gains when subjecting the internal production to international competition and also made it possible for countries to specialize in different branches of industry and production stages. Moreover, by allowing access to the DCs' technology, along with the expansion of exports, it stimulated internal technological development.

Finally, an obligatory reference concerning the divulgation at an academic, institutional and political level is the *World Development Report* 1987 of the WB. With data concerning 41 LDCs, considering two periods of time (1963-1973) and (1973-1985), it grouped the countries in four groups according to the commercial strategy adopted (strongly extroverted, moderately extroverted, moderately introverted and strongly introverted). As a result, it came to the conclusion that the extroverted strategy was superior and decided that the fastest, most sustainable and even most balanced (in terms of personal distribution of income) EG was obtained with this commercial orientation.

1.4 Models of Endogenous Growth and International Trade

In the field of the IT theory, the 'paradox of Leontief' originated debates and controversies leading to the appearance of new developments, which tried to explain the advantages not from the standpoint of a static natural situation but circumscribed to an evolutional process, associated with the EG, where the structural characteristic from which they proceed is continuously under change. The EG theory also suffered significant developments with the models of endogenous EG. These models identify the moving force of growth, its respective dynamics and the forces that influence its accumulation (case of the IT). Thus, these placed the accumulation of human capital and the production and the diffusion of technological innovations in the forefront. The parallelism of these elements with the evolution of the theory of IT isn't accidental. In fact, the models of endogenous EG evolved towards an integrated analysis of the EG and of the IT, recovering in this sense the classical tradition that had been interrupted with the neoclassical separation.

The models of endogenous EG did not come about by accident. Being concerned with the exact microeconomic foundations, they are consequence of the general development of economic theory. We should mention the developments and dissatisfaction with Solow's work, the earlier studies of

themes such as learning by doing [Arrow (1962)], the role of human capital [Uzawa (1965)], increasing returns to scale [Kaldor (1961)] and even the idea of per capita growth sustained by increasing income from the investment in capital goods, which include human capital, dating back to Knight (1944); as well as the inspiration provided by countless authors which have already been mentioned, since Adam Smith.

In accordance with this recent developments, we open the section with a brief and special reference to Lucas' second model (1988) and to the models of endogenous Research and Development (R&D) devised by Romer (1990 and 1993), Grossman and Helpman (1990, 1991a and 1991b) and Aghion and Howitt (1992). We conclude with the mention of several applications.

The Model of Lucas and the Models of Endogenous R&D

In the model of *learning by doing and comparative advantage,* of 1988, Lucas deals with the relation between IT and EG. Essentially, he considered the function of aggregate production with two consumption goods and only one production function, human capital, whose rate of accumulation depended on the quantity of labour connected with production (thus expressing the learning effects). He concluded that with IT each country would specialize in the good for which the autarky donation of human capital presented a comparative advantage. And this specialization tended to be reinforced because the learning took place in the specialized sector. Accordingly, if the rate of learning differed from sector to sector, the rates of EG would be different from country to country.

In the endogenous EG models devised by Romer (1986) and Lucas (1988), the production function of the economy resulted from the aggregation of the firms. Consequently, they turned out to be extremely aggregate and incapable of correctly explaining the microeconomic foundations capable of justifying the functioning of externalities and the agents' investment decisions. A second generation of models [Romer (1990 and 1993), Grossman and Helpman (1990, 1991a and 1991b) and Aghion and Howitt (1992)] considered innovations to be the foundation of the EG process. The innovations were the result of an explicit activity of R&D that occurred in the firms, with the result of R&D being the main determinant of the EG rate.

Technological knowledge is by nature a good without rivalry of use (public good). The market system can't correctly guarantee its production without some public intervention in implementing a system of patents. This system endows technology with the economic nature of a private good, in which the exclusion of use is possible, and which therefore can be sold. An economic problem immediately arises. By definition, the patent places the holder in a monopoly position, and by exploring that position he gains a monopoly rent. On the other hand, the patent entails a fixed cost for the user because its price is generally independent of use. A dilemma of economic policy also subsists in these models, in relation to the diffusion of innovations.

Self-Assessment

1. **Choose the correct options:**
    (*i*) International Trade on Economic Growth were first pointed out by
        (*a*) Smith
        (*b*) Marshall
        (*c*) Ramsey
        (*d*) none of these
    (*ii*) Marginalism   led to a new theory which was called
        (*a*) New Classical
        (*b*) Modern theory
        (*c*) Classical theory
        (*d*) none of these
    (*iii*) The year of birth of the modern neoclassical theory of Economic Growth is
        (*a*) 1959
        (*b*) 1955
        (*c*) 1956
        (*d*) none of these
    (*iv*) The endogenous   economic growth models was devised by
        (*a*) Smith
        (*b*) Romer and Lucas
        (*c*) Englander and Gurney
        (*d*) none of these

1.5 Summary

- In the present work we tried to explain the importance of commercial and technological (dynamic) aspects underlying the IT to EG.

- We noted that the dynamic potential afforded by IT that was pointed out by the classics [Smith (1776)] was disregarded by the 'marginalist revolution'. This was due to the fact that the 'marginalist revolution' studies temporarily left out the lines of the long-term evolution of the economy. As we know, after 1870 the EG was no longer viewed as a great issue for economists due, as it seems, to the perspectives opened by IR. Nevertheless, as exceptions to the rule, authors like Marshall, Young and Schumpeter still dealt with the importance of IT to EG. On the other hand, for instance, the main development in what concerns the scope of the IT theory (the Heckscher-Ohlin-Samuelson model) came to the conclusion that countries benefited from the opening to IT; however, it did no more than identify static gains. But existing studies — for example, Baldwin (1984) — conclude that the static effects (gains only for the increase in the level of per capita income) are very modest.

- It was in this context that, namely after WWII, occurred some reactions to the classical and neoclassical theories which ended up being put to practice in the experiments with introverted and protectionist growth, specially in Latin America. In short, the defenders of these theses maintain that the relevant products as regards IT were produced in keeping with the appeals of the DCs markets and their technologies. Thus, the LDCs were in a disadvantageous situation due to their reduced dimension and sophistication of their markets, as well as to the weak capacity for technological innovation and to the commercial intervention in what concerns the DCs consumers.

- The interest for the EG reawakened, however, with the works of Solow (1956 and 1957). From then on there was a real concern in analyzing the questions belonging to growth in a quantified and systematized way (with a clear distinction between questions belonging to growth and questions belonging to development).

- It should be noted, however, that Solow's (and Swan's) neoclassical growth model assumed technological progress to be exogenous, not because this was a realistic assumption, but because it was the only tractable one. This suggests that interaction with other countries may have no effect on an economy's long term rate of growth. Nevertheless, there may be some interesting effects of openness in the long term level of welfare, and in the transition to the steady state. In the open economy version of the neoclassical model, international flows of capital raise the rate of convergence to the steady state.

- In the late 1950s, the seminal paper by Solow (1957) attempted to account for economic growth in the US, finding it to be not fully explained by the increase in productive inputs such as labour and capital alone. The largest part of growth was thus attributed to a residual. In subsequent research, much effort was devoted to trying to better understand the origin of productivity increases by squeezing down the residual, by introducing other variables such as accumulation of human capital, economies of scale, a better allocation of resources and new generations of more productive machines. However, even with the introduction of new variables an unexplained residual remained.

- Therefore, on the one hand, the attempt to determine sources of growth in their entirety and, on the other hand, the failure of introverted growth experiences and the association of fast EG to the opening of IT and to the resulting international specialisation in several countries led to the undertaking of research on trade and growth (which adopted the neoclassical framework). We mentioned some theoretical studies — structuralist syntheses, analyses that underscore economic integration, the models of Findlay (1980 and 1984) and Feder (1982) —, empirical applications— among others, structuralist studies, Feder (1982) and Ram (1987) — and studies and/or recommendations about the external commercial policy — among others, UN recommendations, Balassa (1978, 1986 and 1987), Krueger (1985) and WB (1987) — whose defining characteristic is to view IT (above all the exporting component) as an explanatory variable of EG.

- Generally, they associate that situation with a better allocation of resources (according to the comparative advantages), with a greater utilisation of the productive capacity (which makes it possible to obtain economies of scale), with the greater propensity to implement technological improvement (in answering to the greater competition that they face) and with the higher level of employment created in comparison with introverted strategies.

- Although this body of literature enlarged the original framework, technology was still treated as a public good.

- However, on the one hand, in view of the neoclassical theory's limitations (mainly because the technological progress is exogenous but also because, in open economies, this suggests that, in practice, the increase of the convergence among countries is not verifiable) and, on the other hand, in view of the many developments and suggestions which are afforded by Smith, Schumpeter, Knight, Arrow, Kaldor and Uzawa, among others, economists have recently started to model the process of knowledge accumulation, and the resulting literature is known as endogenous growth theory. This allows us to develop tractable and flexible models that embody the vision of economics life as an endless succession of innovation and change wrought by competition.

- These growth models allow for an economy to be able to reach a balanced growth path through endogenous forces and underscore the microeconomic foundations of the growth process, identifying in detail the driving force of growth (which is knowledge, generally under the form of technological innovation), its respective dynamics as well as the driving forces which influence its accumulation. Thus, in most new models the determining factor of economic growth is endogenous innovation, and this innovation is still influenced by IT. Consequently, the modelling which these new models afford brought with it a more exact approach to the relation between EG and IT. So we can say that the dynamic potential created by IT was decisively recovered more recently with the advent of the models of endogenous growth.

- Furthermore, the endogenous approach, bringing increasing returns and non-competitive market structures into the core of growth analysis, made it so that perfect competition would no longer be a *sine qua non* condition for optimal trajectories of growth to exist. The growth path may not be optimal. So, the governmental intervention may be useful in order to move the growth path towards the optimal one.

- Regarding the contribution of IT to EG, in light of the new approach, we alluded to Romer's work (1990), which viewed IT as a motivating factor of growth, when integrating economies with different levels of human capital. We also saw that the assumptions as to differences among countries condition trade patterns and their effect on growth. With respect to this, Lucas (1988) and Grossman and Helpman (1991a) assume that the only differences among countries have to do with initial provision of factors, whereas Grossman and Helpman (1990) point to differences in respect to the countries' technological capacities.

- The works of Grossman and Helpman (1991b and 1991c) and Rivera-Batiz and Romer (1991a) have also helped clarify why a country's participation in an integrated world economy can speed up its growth : among other reasons, it allows access to a wider base of technological knowledge, it makes technological diffusion easier, it motivates research and avoids redundancies in research. We also presented Romer's work (1993), which recommended that the LDCs open to the foreign investment with more advanced technology so that they could register increases in the rate of innovation and in the economy's rate of growth.

- In this context, the abundant empirical evidence, specifically, suggests that trade openness tends to be beneficial for growth. Especially for the DCs, because they affect the domestic rates of innovation. And for the LDCs (which hardly invest in R&D) because of the dynamic effects of the economic integration with DCs, the catch-up of the convergence, the importation of capital goods and the capacity for adaptation and implementation of innovations. Finally, let us mention that the intensity of dynamic effects depends simultaneously on the geographic structure of international trade (*i.e.*, on the level of development of trade partners), on the composition and

intensity of IT and on the capacity for internal technological adaptation, which is made possible through higher levels of human capital, as suggested, for example, by Lucas (1988) and Romer (1990).

1.6 Key-Words

1. Neoclassical theory	:	An economic theory that outlines how a steady economic growth rate will be accomplished with the proper amounts of the three driving forces: labor, capital and technology. The theory states that by varying the amounts of labor and capital in the production function, an equilibrium state can be accomplished. When a new technology becomes available, the labor and capital need to be adjusted to maintain growth equilibrium.
2. Endogenous growth theory	:	This theory holds that economic growth is primarily the result of endogenous and not external forces. Endogenous growth theory holds that investment in human capital, innovation, and knowledge are significant contributors to economic growth. The theory also focuses on positive externalities and spillover effects of a knowledge-based economy which will lead to economic development. The endogenous growth theory also holds that policy measures can have an impact on the long-run growth rate of an economy. For example, subsidies for research and development or education increase the growth rate in some endogenous growth models by increasing the incentive for innovation.

1.7 Review Questions

1. Explain trade as an Engine of growth.
2. What do you mean by Post classical period? Discuss international growth and trade in context of post classical period.
3. What are the models of endogenous growth and international growth? Discuss.

Answers: Self-Assessment

1. (*i*) (*a*)   (*ii*) (*a*)   (*iii*) (*c*)   (*iv*) (*b*)

1.8 Further Readings

1. Abramowitz, M. (1956), Resource and output trends in the United States since 1870, *American Economic Review,* 46(2), May, pp. 5-23.

2. Aghion, P. and Howitt, P. (1992), A model of growth through creative destruction, *Econometrica,* vol. 60, no 2, pp. 323-351.

3. Backhouse, Roger E. (1985) — *A History of Modern Economic Analysis*— Oxford : Basil Blackwell.

4. Balassa, B. (1963), An empirical demonstration of classical comparative cost theory, *Review of Economics and Statistics,* vol. 45, pp. 231-238.

5. Balassa, B. (1978), Exports and economic growth : further evidence, *Journal of Development Economics,* vol. 5, no 2, pp. 181-189.

6. Balassa, B. (1986), Policy responses to exogenous shocks in developing countries, *American Economic Review,* Vol. 76, no 2, May, pp. 75-78.

Unit 2 : Measurement of Gains from Trade

Objectives

After reading this Unit students will be able to:

•   Describe how to measure Gains from Trade.

•   Explain an Application to Fair Allocation.

Introduction

In the classical model of exchange gains from trade can be obtained if there exists a feasible allocation which each agent prefers to her endowment. But, how can we measure the gains from trade in an economy ? For interpersonal comparisons of welfare are typically not meaningful, we propose measuring gains from trade in terms of quantities of goods, avoiding welfare comparisons. To do so, we search for a "reference allocation", composed of "reference bundles", one for each agent, such that : (i) each agent is indifferent between her endowment and her reference bundle, and (ii) the reference allocation is feasible. Since no welfare gains are achieved, the difference between the aggregate endowment and the resources at the reference allocation provides a measure of the gains from trading from the endowment profile to the reference allocation.

In this manner, we obtain a measure of gains from trade in terms of quantities of goods. But, for most economies, there is a continuum of reference allocations and the vectors of resources saved by trading to each differ. The set of all such vectors defines the set of possible gains from trade of the economy. We introduce the notion of a (vector-valued) "metric", to select, for each economy, one representative vector from its set of possible gains from trade.

Two economies may differ in preferences and endowment profiles but have equal sets of possible gains from trade. Our premise is that, if two economies have equal sets of possible gains from trade, a metric should not distinguish between them, and should select the same representative vector of gains from trade in both economies. Thus, the notion of a metric is similar to the notion of a solution for bargaining problems. A "bargaining problem" consists of a set of utility profiles and a disagreement point; a "solution" maps each bargaining problem into a utility profile. In our setting, the set of utility profiles corresponds to the set of possible gains from trade, and the disagreement point corresponds to each agent consuming her endowment and no resources being saved. A metric maps each set of possible gains from trade into a vector of gains from trade.

We follow an approach used in bargaining theory and look for metrics satisfying certain desirable properties. Metrics should select a vector representative of the size of the set of possible gains from trade. Thus, we look for metrics satisfying the following three properties that have strong intuitive

appeal for our setting. The first property is "maximality" : for each economy, no reference allocation leads to a larger vector of gains from trade than the vector selected by the metric. The second property is "monotonicity" with respect to set inclusion : a metric selects a larger vector of gains from trade in an economy with a larger set of possible gains from trade. The third property is "homogeneity" : a homogeneous expansion of the set of possible gains from trade leads to a homogeneous expansion of the vector selected by the metric.

2.1 Gains from Trade

The combination of consumer surplus and producer surplus obtained by buyers and sellers when engaging in a market exchange. Gains from trade arise because buyers are typically willing and able to pay a higher price to purchase a good than what they end up paying and because sellers are typically willing and able to accept a lower price to sell a good than what they end up receiving. Both sides of the market exchange are thus better off, have a net gain in welfare, by making the trade. While all types of market exchanges generate gains from trade, this topic is perhaps most important for an understanding of international trade.

> *Did u know?* Buyers and sellers engage in market exchanges because they benefit from the trade.

As a generally rule both sides are better off after the exchange than they were before the exchange. Buyers are better off because they have a net gain in consumer surplus. Sellers are better off because they have a net gain in producer surplus.

Voluntary market exchanges are undertaken because they are beneficial to both sides of the transaction. If buyers and sellers did not gain from the trade, then they would not voluntarily undertake the trade.

While the gains obtained from market exchanges provides insight into all forms of trading and the very existence of a market-based economy used to allocate resources, it also provides a great deal of insight into trading among nations, that is, international trade. When two nations engage in trade they do so because they gain from the trade. Both countries are better off after the trade than they were before.

Market Trades

The motivation behind international trade is essentially the same as for any market exchange. People buy and sell goods because they expect to be better off after the exchange than they were before. To illustrate this, consider the motivation of two hypothetical people -- Horst Duncanstein and Francine von Sutter -- who are primed to do a little exchanging.

• **From the Buying Side:** First, consider the situation facing Horst Duncanstein, who is exceptionally fond of turnip lasagna. Eating turnip lasagna makes Horst a happy fellow. It improves his level of well being. It satisfies his wants and needs.

To this end, Horst is willing to pay a price for the turnips needed to make his turnip lasagna. Horst has a maximum price that he is willing to pay for the needed turnips -- a demand price. If the price is too high, then he will not purchase turnips, opting to consume another good, perhaps carrots to be used in a carrot casserole. However, should the price he pays for his turnips be less than his demand price, then he comes out ahead. He pays less than the value he receives, what is termed consumer surplus. He gains from this trade.

• **From the Selling Side:** Second, consider the situation facing Francine von Sutter, a turnip farmer. While Francine does not have a particular fondness for turnips, she does enjoy the farming business. She has the land, labor, and capital needed to produce turnips.

To this end, Francine is willing to provide turnips to willing buyers so long as she can cover the cost of production. Francine has a minimum price that she is willing to accept to produce turnips -- a supply price. If the price is too low, then she will not produce turnips, opting to produce another

good, perhaps carrots. However, should the price she receives for her turnips be greater than her supply price, then she comes out ahead. She receives more than the cost of production incurred, what is termed producer surplus. She gains from this trade.

Putting Horst and Francine together is bound to be beneficial for both. If Horst pays less than his demand price, then he gains from the trade. If Francine receives more than her supply price, then she also gains from this trade. It is a win-win exchange.

In the extreme case, it is possible that the price Horst pays is exactly his demand price or the price Francine receives is exactly her supply price. In this case, one side or the other does not gain from the trade, but neither does that side lose.

However, should the price rise above the maximum demand price Horst is willing to pay or fall below the minimum supply price Francine is willing to pay, then the exchange will not occur. One side or the other will opt out of the trade.

The end result of such voluntary trades between buyers like Horst and sellers like Francine is that one side or the other, and usually both, gain from the trade. If they did not gain (or at least break even), then they would not voluntarily engage in the exchange.

Graphical Gains

The gains obtained from market exchanges can be illustrated using the exhibit to the right. This exhibit presents a standard market graph. The negatively-sloped demand curve, D, represents the demand price that buyers (like Horst) are willing and able to pay to purchase different quantities of turnips. The positively-sloped supply curve, S, represents the supply price that sellers (like Francine) are willing and able to accept to sell different quantities of turnips.

If this is a competitive market, free of other market failures and other annoying complications, then the intersection of the demand and supply curves gives rise to the equilibrium price and equilibrium quantity. The relation between the market equilibrium price, the demand price on the demand curve, and the supply on the supply curve indicates the gains from trade.

The area above the equilibrium price and below the demand curve is the consumer surplus generated by this market. The area below the equilibrium price and above the supply curve is the producer surplus generated by this market.

The combination of these two areas, the area above the supply curve and below the demand curve, is the gains from trade generated by this market. This is extra satisfaction, welfare, profit, etc. that would not exist if this market exchange did not take place.

Gaining from International Trades

The only difference between regular market trades, such as that between Horst and Francine, and international trades is the location of the buyers and sellers. If Horst lives in one nation, such as the hypothetical Republic of Northwest Queoldiola, and Francine lives in another, such as the equally hypothetical United Provinces of Csonda, then the previous market exchange example is also an international trade. But the gains from trade still result.

As a matter of fact, Horst does live in the Republic of Northwest Queoldiola and like other Queoldiolan's, he loves his turnip lasagna. And Francine is a turnip-growing citizen of the United Provinces of Csonda, and she is eager to sell her product to buyers from other lands.

Horst and Francine gain from this turnip exchange, but so too do their home nations. Northwest Queoldiola ends up with a bit more consumer surplus, thanks to that obtained by Horst, and the Csonda ends up with bit more producer surplus, thanks to that obtained by Francine.
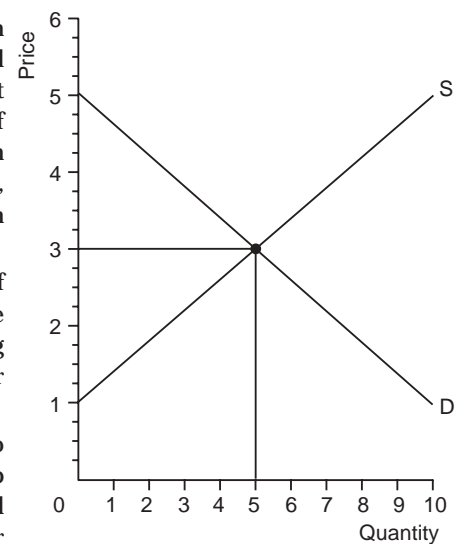

Figure 2.1

2.2 How to Measure Gains from Trade

Our proposal for measuring gains from trade in terms of quantities of goods can be interpreted as a generalization, to a multi-agent setting, of some existing measures of welfare changes in single agent decision making settings. The equivalent variation and compensating variation are measures of welfare changes in terms of the difference in expenditure required to keep an agent's welfare unchanged after a change in prices. In our setting, change does not come from prices but from trading among agents, and measuring gains in terms of quantities of goods seems natural in the absence of pre-specified prices. In settings of choice under uncertainty, the risk premium measures how much an agent is willing to forgo in order to obtain a constant consumption stream; the certainty equivalent measures the level of constant consumption across states that leaves the agent's welfare unchanged. In our setting, we measure how much a set of agents can gain by redistributing risk among them.

> **Notes**
> Measuring gains from trade is equivalent to measuring the inefficiency of the endowment. A measure the inefficiency of an allocation (or of the endowment profile) is its "coefficient of resource utilization" (Debreu 1951). It assigns to each maximal vector in the set of possible gains from trade a number equal to the dot product of the vector and its supporting price. Then, it measures the inefficiency of the allocation by the maximal such value.

This way of measuring the inefficiency of an allocation is similar to ours. It also considers the set of possible gains from trade of the economy. But, instead of measuring gains from trade by a vector of commodities, it measures gains from trade by a scalar. Using a real-valued metric implies that we can order the set of all economies according to their gains from trade. Using a vector-valued metric allows for a partial order, which may be desirable if differences in goods require asymmetric treatment across them. Moreover, this measure is not monotonic, an increase in the set of possible gains from trade can lead to a decrease in the measurement of this gains.

Another advantage of using a vector-valued metric over a real-valued one, is that a vector-valued metric leads to a natural allocation at which gains from trade are distributed fairly. The theory of fair allocation can be categorized according to the nature of the problem under study : First, situations where a social endowment has to be divided among a set of agents. Second, situations where agents have private endowments and redistribution (trading) is possible. For the problem of allocating a social endowment two notions of fairness are prominent. First is no-envy (Foley 1967) : no agent should prefer another agent's bundle over her own (see Kolm (1998) and Varian (1976)). Second is egalitarian equivalence (Panzer and Schmeidler 1978) : there exists a reference bundle such that each agent is indifferent between her bundle and the reference bundle. For the problem of redistributing individual endowments these two notions can be adapted. No-envy in trades states that no agent prefers another agent's trade over her own. Egalitarian-equivalence from endowments states that there exists a reference vector such that each agent is indifferent between her bundle and the bundle obtained from the sum of her endowment and the reference vector.

Recently, a notion similar to egalitarian equivalence was proposed for economies with individual endowments : an allocation is fair if it is welfare equivalent to an allocation obtained from summing to the endowment profile a vector of fair "concessions" (Pérez-Castrillo and Wettstein 2006). This notion generalizes egalitarian equivalence in two ways : first, it allows for differences in the reference bundles according to differences in individual endowments; second, it allows for differences in concessions.

Our notion of fairness is similar to Pérez-Castrillo and Wettstein (2006) but it differs in two ways. First, our reference allocation is welfare equivalent to the endowment profile, and we sum to the reference allocation the vector of contributions. Second, our vector of contributions differs from their vector of concessions. Also, our results differ in form from theirs. They show existence of fair and efficient allocations; we do not obtain a fair and efficient allocation immediately, but propose a recursive

procedure which is fair at each step, and obtains an efficient allocation at the limit. Also, we provide an algorithm to reach it.

2.3 Measurements of Gain from International Trade

The gains from international trade are measurable. Prof. Jacob Viner says that the classical economists had adopted three methods of measuring the gains from internatilnal trade.

1. Measurement of increased real income by comparative cost theory.
2. Increase in the level of national income.
3. Improvement in terms of trade.

However, with the introduction of J.S. Mill's theory of Reciprocal Demand, the most frequently used method of measuring gains from trade is the terms of trade method.

Hence, in order to measure gains from trade, three approaches are used.

1. Ricardo's Approach
2. J.S. Mills Ajpproach
3. Modern Approach or Samuelson's Approach

Ricardo's Approach

According to Ricardo, a country would export those goods in which its comparative cost of production is less. This is proved in an example and diagram.

> India: 15 units of cotton or 20 units of wheat
> Pakistan: 10 units of cotton or 10 units of wheat

(*i*)

If in every country only two units of the factor are used, the product would be as under:

> India: 15 units of cotton + 20 units of wheat
> Pakistan: 10 units of cotton + 10 units of wheat

(*ii*)

1. Without specialization, if both the countries produce both the commodities, the total production in the two countries would be as under:

> India + Pakistan = 25 units of cotton + 30 units of wheat
> (25C + 30W)

(*iii*)

2. If there is specialization on the basis of comparative cost theory, in India specializing in the production of cotton, total production would be:

> India = 25 units of wheat
> Pakistan = 20 units of cotton

(*iv*)

India + Pakistan = 40 units of wheat + 20 units of cotton

Comparing situations (*ii*) and (*iii*), reveals that due to specialization in the two countries, production of wheat increases by 10 units whereas there is loss of 5 units of cotton.

> Net Result = + 10 units W − 5 units C

(*v*)

From situation (*i*) it is revealed that 10 units of wheat in India is equal to 7.5 units of cotton and in Pakistan it is equal to 10 units of cotton i.e.

10 W = 7.5 C or 10 W = 10 C

So the net result is:

+ 7.5 C – 5 C

or   10 C – 5 C

i.e.   + 2.5 C or + 5 Cotton

Thus, specialization results in net gain of 2.5 or 5 units of cotton which is distributed between the two countries.

The gain that a country enjoys by pursuing trade according to this theory is illustrated by the following Figure 2.2.

In Figure 2.2 X-commodity is shown on OX-axis and Y-commodity on OY-axis. Suppose, in case of trade, AB is the production possibility curve that indicates different combinations of X-commodity and Y-commodity produced by the given number of labour. Point 'E' on AB curve indicates equilibrium position of the country. After entering into trade, the production possibility curve shifts and assumes the shape of BC curve. Slope of BC cirve indicates international price ration of the country. Suppose this country is in equilibrium at point 'F' on ABI curve. If this country produces a combination of X-commodity and Y-commodity as shown by point 'F', it will have to increase the number of labourers to such an extent that domestic production possibility curve shifts from AB to A_1B_1.
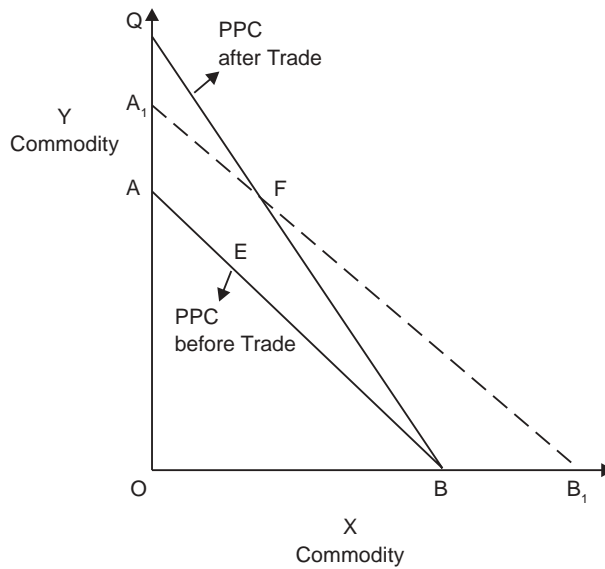
Figure 2.2: Gains from Trade

Thus the amount of gain from the trade will be measured by BB_1/OB.

Criticism

The main points of criticism of gain from international trade occurred as a result of comparative cost or Richardo's teory are as under:

1. According to later economists, Ricardo has unnecessarily exaggerated the gain from international trade. Ricardo's theory does not apply to those countries which cannot produce the imported goods or can produce the same only at higher cost.

2. Mill feels that Ricardo's theory does explain the reason why international trade takes place but it does not explain the quantum of gain and how the same is distributed among different countries.

Mill's Approach

J.S. Mill analyzed the gain as well as the distribution of the gain from international trade in terms of his theory of reciprocal demand. According to Mill, it is the reciprocal demand that determines terms of trade which, in turn, determines the distribution of gains from trade of each country. The term 'terms of trade' refers to the batter terms of trade between the two countries, i.e., the ration of the quantity of imports for a given quantity of exports of a country.

To take an example; in country A, 2 units of labour produce 10 units of X and 10 units of Y, while in country B the same labour produces 6X and 8Y. The domestic exchange ratio (or domestic terms of trade) in country A is 1X = 1Y, and in country B, 1X = 1.33Y. This means that one unit of X can be exchanged with one unit of Y in country A or 1.33 units of Y in country B. Thus, the termsof trade between the two countries will lie between IX or 1Y or 1.33 Y.
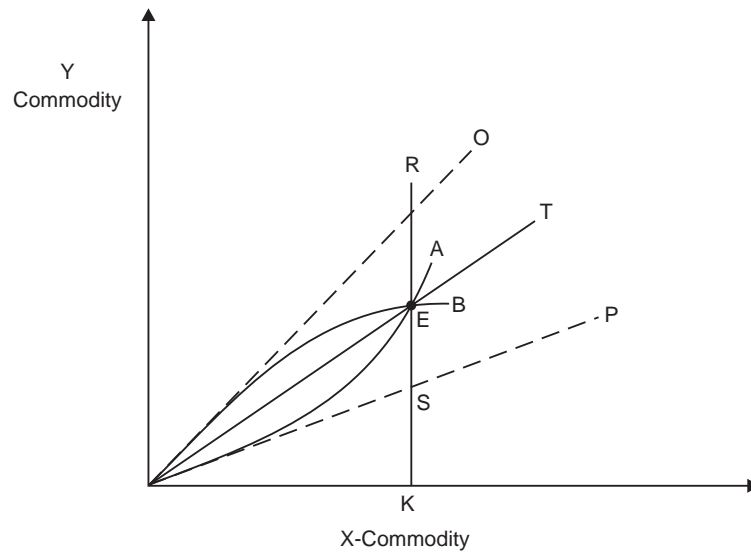
Figure 2.3: Marshall-Edge worth offer Curves and Distribution of Gains from Trade

However, the actual exchange ratio will depend upon the reciprocal demand, i.e., "the relative strength and elasticity of demand of the two trading countries for each other's product in terms of their own product." If A's demand for commodity Y is more intense (inelastic), then the terms of trade will be nearer 1X = 1Y. The terms of trade will move in favour of B and against country A. B will gain more and A less. On the other hand, if A's demand for commodity Y is less intense (more elastic), then the terms of trade will be nearer 1X = 1.33 Y. The terms of trade will move in favour of A and against B. A will gain more, and B less.

The distribution of gains from trade is explained K in terms of the Marshall-Edgeworth offers curves in Figure 2.3 OA is the offer curve of country A, and OB of country B. OP and OQ are the domestic constant cost ratios of producing X and Y in country A and B respectively. These rays are, in fact, the limits within which the terms of trade between the two countries lie. However, the actual terms of trade are settled at E the point of intersection of OA and OB. The line OT represents the equilibrium terms of trade at E.

The cost ratio within country A is KS units of Y and OK units of X. But it gets KE units of Y through trade. SE units of Y is, therefore, its gain. The cost ratio within country B is KR units of Y and OK units of X. But it imports OK units of X from country A in exchange for only KE units of Y. ER units of Y is its gain. Thus, both countries ain by entering into trade.

Haberler's Proof of the Gains of Trade

Haberler has specified the gains of trade in a given diagram. In Figure 2.4 AA is the production possibility curve. Before trade H is the equilibrium point showing the state of production and consumption. The slope of the tangent DD at H shows the price ratio before trade. After international trade price ratio is shown by PP line which is tangent at point T on the PPC. Point T represents production equilibrium point and H' represents competition equilibrium point. At H' country exports H'L quantity of X and import I.T quantity of Y commodity.
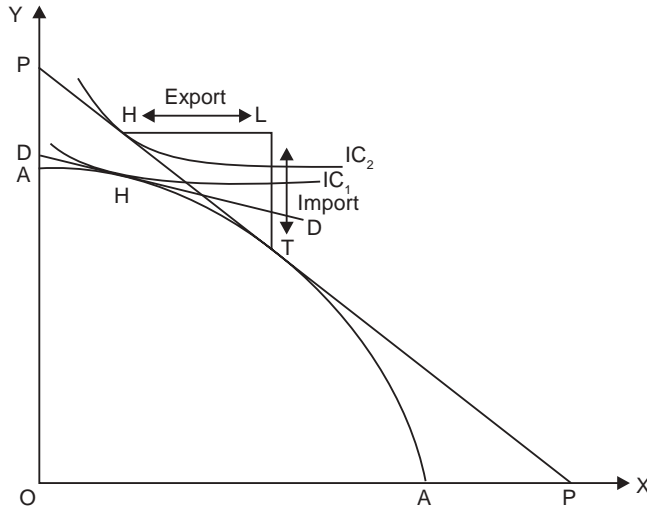
Figure 2.4: Gains from Trade - Haberler's Approach

On the basis of community IC it has been clarified that point H' is superior to H as at H' higher indifference curve is tangent at H' than at H. Hence these are gains from international trade. It should however, be clarified that since Haberler was not in favour of using community indifference curves, point H1 can prove to be superior to H if H1 is above and to the right of H. In this case at H1 international trade causes rise in the quantity of both X and Y goods. This is nothing but gain from trade.

Modern Approach (Samuelson's Approach)

In modern trade theory, the gains from international trade are clearly differentiated between the gains from exchange and the gains from specialization. The analysis is explained in terms of the general equilibrium of a closed economy by taking demand and supply. It is characterized by the tangency of a community indifference curve with the transformation curve, and the equality of the marginal rates of substitution between commodities in consumption and production with the domestic terms of trade or commodity price ratio. "*The introduction of international trade permiss the realization of a gain from exchange and gan from specialization. When equilibrium is essablished and these gains are maximized, the new marginal rate of transformations in production and the new marginal rate of substitution in consumption are equal to the international price ratio or terms of trade.*" Thus, both producers and consumers gain from international trade by producing and consuming more than the pre-trade level.
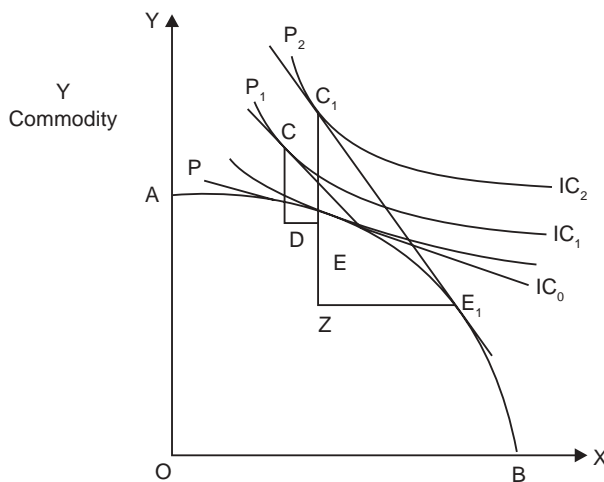
Figure 2.5: Gains from Trade-Samuelson's Approach

Figure 2.5 explains the gains from international trade. AB is the transformation curve representing the supply side and IC_0 is the community indifference curve representing the demand side of an economy. The closed economy (no trade) equilibrium is shown by point E where the AB and IC_0 curves are tangent to each other and both equal the domestic terms of trade or commodity price ratio (line) PP.

With the introduction of international (or free) trade the international price ratio (terms of trade) will be different from the domestic price ratio (terms of trade). It is shown as P_1 and is steeper than the domestic price ratio P. It means that the price of commodity X has increased in relation to commodity Y in the world market. At the international price line P_1, the consumers move to point C on a higher community indifference curve CI, from point E on the IC_0 curve. This movement from E to C measures the gan from exchange or consumption gain wth no change in production.

Since the price of X has increased in the world marker, producers increase its production and decrease that of Y. This lead to movement along the transformation curve from point E to E_1 where the international price line P_2 is tangent equals the international price ratio. The new world terms of trade ratio P_2 is the same as P_1 because it is parallel to P_1. At E_1 the country exports E_1Z of X in exchange for ZC_1 imports of Y.

As a result of increased specialization in the production of X, there is shift in consumption from point C on the IC_1 curve to point C' on the IC_2 curve, where consumers consume larger quantities of both X and Y. This movement from C to C' measures the gain from specialization in production or production gain. At C', the marginal rate of substitution and the international price ratio are equal. Hence the gains from international trade are maximized at points E_1 and C' because the marginal rate of transformation in production and the marginal rate of substitution in consumption are equal to the international price ratio P_2. The total gain from free tradeis the sum of the consumption are equal to the international price ratio P_2. The total gain from free trade is the sum of the consumption and production gains and is shown as improvement in welfare from IC_0 to IC_2.

2.4 An Application to Fair Allocation

Once we have a measure of the gains from trade in an economy, how can we distribute them fairly? An "allocation rule" recommends for each economy a set of feasible allocations. We look for allocations rules that distribute gains from trade fairly. First, we propose a method to determine the contribution of each agent to the gains from trade. Then, we declare an allocation "fair" if each agent obtains her contribution to the gains from trade. Finally, we propose an allocation rule which assigns to each agent her contribution to the gains from trade. This rule is not efficient, but we show that a recursive procedure distributes gains from trade fairly at each step and defines an efficient rule.

2.4.1 Contributions to Gains from Trade

In order to determine each agent's contribution to the gains from trade, we propose to use the solution concept of the theory of cooperative games known as the Shapley-value, using its interpretation as rewarding agents as a function of their "marginal contributions" to all subgroups. We measure each agent's contribution to the gains from trade as the "marginal gains" in each subpopulation.

First, we generalize the definition of the *weighted-gains family* to allow for variable populations. For each subpopulation $N' \subset N$ and each economy $(R, \omega) \in \varepsilon$, the α -*weighted-gains metric* measures gains from trade of the subeconomy $(R_{N'}, \omega_{N'})$ by the largest vector $z \in G(R_{N'}, \omega_{N'})$ proportional to α.

Self -Assessment

1. **Choose the correct option:**

    (*i*) Mercantilism advocated that a country

    (*a*) prohibit all exports because it viewed trade as zero sum, believing that one nation's gain was another nation's loss.

    (*b*) Should pursue Free trade because it viewed trade as zero sum, and therefore, that trade would always be balanced.

   (*c*) promote exports over imports because it viewed trade as zero sum, believing that one nation's gain was another nation's loss.

   (*d*) use tariffs rather than quotas. Because mercantilism views trade as zero sum, transparent trade barriers are viewed as best.

 (*ii*) Adam Smith was critical of trade barriers, since he believed that trade barriers

   (*a*) Are a good and appropriate source of government revenue.

   (*b*) Reduce specialization, technological progress and wealth creation.

   (*c*) Harm a country's relations with its colonies.

   (*d*) All of the above.

(*iii*) The Resource Curse refers to the idea that

   (*a*) Countries that focus on a resource, such as oil, may suffer from macroeconomic instability.

   (*b*) Countries that don't have resources, such as oil, are at the mercy of resource-rich countries.

   (*c*) Countries that don't have resources are more likely to have corrupt governments.

   (*d*) Countries that don't have resources may suffer from macroeconomic instability.

(*iv*) Competitive advantage and comparative advantage will differ for China if

   (*a*) True production costs are inaccurately measured due to production externalities such as pollution.

   (*b*) None of the above. Competitive advantage and comparative advantage are always the same.

   (*c*) The Chinese currency (the Renimbi) is overvalued.

   (*d*) Both A and B.

 (*v*) When a country trades according to principles of comparative advantage.

   (*a*) Some workers will be hurt due to dislocation. Their loses are so large that it is impossible to offset these losses through the sharing of the gains from trade.

   (*b*) All workers will lose due to foreign competition.

   (*c*) All workers will gain.

   (*d*) Some workers will be hurt due to dislocation, though their losses could be offset if the gains from trade were shared.

2.5 Summary

- We proposed a method to measure gains from trade. We avoided interpersonal comparisons of welfare by denning gains in terms of quantities of goods. To do so, we introduced the notion of a *metric*. A *metric* measures gains from trade by a vector of quantities of goods which can be saved while keeping each agent's welfare unaffected. We characterized the family of *metrics* satisfying some intuitive properties (Theorem 1). This method of measuring gains is applicable to a wide variety of settings. It can be interpreted as generalizations of existing measures of welfare changes in single agent settings to multi-agent settings.

- Then, we proposed an application to fair allocation. Based on Shapley's algorithm, we obtained a way of measuring each agent's contribution to the gains from trade. We declared an allocation fair if each agent receives her contribution to the gains from trade. We defined a fair allocation rule that assigns to each agent her contribution. This rule is inefficient, but we show that a recursive procedure, which is fair at each step of the recursion, yields an efficient rule (Theorem 2).

- Now, we discuss relaxing some of the assumptions. First, we discuss relaxing the assumptions of boundary aversion and strict monotonicity of preferences. Then, we discuss relaxing some of the properties on *metrics.*

- Throughout the paper, we assumed that preferences are strictly monotonic and satisfy boundary aversion. When preferences fail either of these properties but are (weakly) monotonic, Proposition 1 no longer holds. The *sets of possible gains from trade* of some economies are not strictly

comprehensive; but they are still closed, convex, bounded, and comprehensive. Proposition 2 still holds. Moreover, for each closed, convex, bounded, and comprehensive set we can find an economy whose set of possible gains from trade and this set coincide.

- The domain of closed, convex, bounded, and comprehensive sets is the usual domain of problems in bargaining theory. It is well-known that on this domain there is no *maximal* and *monotonic* solution. We can weaken *monotonicity* to hold whenever the smaller of the two sets of gains from trade is strictly comprehensive, and obtain a generalized version of the *weighted-gains* family. A member of this generalized family measures gains from trade by the largest vector proportional to a vector of weights, but, if this vector is not maximal, it drops some goods, and continues measuring gains from trade proportional to a restricted vector of weights. We refer to Thomson (2004) for a detailed treatment of this family in the context of bargaining theory.

- For the application to fair allocation, *monotonicity* of the *metric* was necessary for the proof of Theorem 2. As stated in the text, we conjecture that an alternative proof can be obtained without *monotonicity* if we require a welfare improving property.

- Finally, we discuss relaxing the requirement that *metrics* measure equal gains from trade in economies with equal *sets of possible gains from trade.* Sets of possible gains from trade depend on relatively little information about preferences. This property may be desirable when obtaining information is costly, but we may lose too much information in the aggregation procedure. Relaxing this property is an interesting an open question left for future research. For now, we note that *monotonicity* of a *metric* implies this property.

2.6 Key-Words

1. Symmetric : Symmetry generally conveys two primary meanings. The first is an imprecise sense of harmonious or aesthetically pleasing proportionality and balance; such that it reflects beauty or perfection. The second meaning is a precise and well-defined concept of balance or "patterned self-similarity" that can be demonstrated or proved according to the rules of a formal system: by geometry, through physics or otherwise. Although the meanings are distinguishable in some contexts, both meanings of "symmetry" are related and discussed in parallel.

2. Maximality : Of relating to, or consisting of a maximum, being the greatest or highest possible

2.7 Review Questions

1. Write a short note on the gains from trade.
2. Discuss the measurement of gains from trade.
3. What are the concepts of trade? Discuss.

Answers: Self-Assessment

1. (*i*) (*c*) (*ii*) (*b*) (*iii*) (*a*) (*iv*) (*d*) (*v*) (*d*)

2.8 Further Readings

Books

1. **DEBREU, G. (1951) :** "The coefficient of resource utilization," *Econometrica,* 19, 273–292.

2. **DOMINGUEZ, D. (2006) :** "Lower bounds and recursive methods for the problem of adjudicating conflicting claims," *mimeo.*

3. **FOLEY, D. (1967) :** "Resource allocation and the public sector," *Yale Economic Essays,* 7, 45–98.

4. **KALAI, E. (1977) :** "Proportional solutions to bargaining situations : interpersonal utility comparisons," *Econometrica,* 45, 1623–1630.

Unit 3 : Free Trade Theory — Absolute Advantage, Comparative Advantage and Opportunity Cost

---

Objectives

After reading this Unit students will be able to:

- Describe Absolute Advantage Model of Adam Smith.

- Explain Comparative Advantage Model of David Ricardo.

- Understand the Opportunity Cost and the Pure Theory of Trade.

Introduction

The theory of trade has a central place in economic analysis, and underpins the doctrine of free trade. Free trade doctrines have a long and fascinating history in Europe. In 1846 Britain repealed the Corn Laws, an historic event which marked the start of the era of free international trade, and lasted until the great depression of the 1870s. The Corn Laws were the duties on imports of grain, which had been in force in England since the middle of the fifteenth century. Other European countries had similar taxes : France, Sweden, Bavaria, Belgium and Holland.

The reasoning behind the Corn Laws was as follows. Grain, chiefly wheat, is a staple foodstuff, especially important in the diets of labouring people. But its price varies greatly from year to year, depending on the size and quality of harvests. Duties on imports were levied on a sliding scale in order to stabilise the price of wheat. When the domestic price was high because of a poor harvest, duties were lowered to permit imports. When the domestic price was low because of a bumper harvest, import duties were raised.

In the decades leading up to the repeal of the Corn Laws in Britain, the system had fallen into disrepute. In fact the sliding scale of duties was tending to increase rather than reduce fluctuations in the price of wheat. When the domestic price was high, traders tended to withhold supply to raise the price even further. They anticipated that import duties would soon be lowered, which was in fact what tended to happen. Then, when duties fell, traders began to import large quantities of grain. As supply rapidly increased, and prices fell dramatically, import duties were quickly increased. The net effect was to amplify market fluctuations through speculation, making a vulnerable market even more unstable, much to the detriment of consumers.

The Corn Laws had another important effect. They benefited agricultural interests at the expense of the newly emerging manufacturing sectors. High prices of grain, maintained through restricting foreign supply, increased the value of land. Landowners, understandably, came to constitute an important pressure group for the maintenance of the Corn Laws. Against these landed interests were

ranged the burgeoning manufacturing classes. In Britain, the opposition to the Corn Laws centred on Manchester, the home of the textile industry. The 'free traders' as they were called, believed that lower grain prices were needed so that the labouring classes in industrial areas would have access to cheap foodstuffs. Led by Cobden, formerly a manufacturer, the free traders argued for the opening-up of British markets to cheap grain imports from overseas. Manufacturers were also anxious that free trade principles should be reciprocated in other countries, so that foreign markets would be opened up to exports of cheap manufactured goods from Britain.

In Britain free trade principles eventually triumphed. In the twentieth century, with the important exception of the period 1918 to 1939, free trade principles also came to dominate the world economy. In this chapter we explore the economic principles which underpinned the doctrine of free trade, a doctrine which is arguably one of the most robust of any in present-day economics. These principles were reinterpreted in terms of modern economics by the economist Haberler in the 1930s.

Finally, a word of warning — the theory of comparative cost, on which everything in this chapter rests, is deceptively simple! In 1996, the world-famous US economist Paul Krugman came to Manchester, UK, to give a paper to mark the 150 years which had elapsed since the repeal of the Corn Laws. He entitled his address 'Ricardo's Difficult Idea : Why Intellectuals Don't Understand Comparative Advantage'. In it he made clear that intelligent people who read, and even those who write about world trade, often fail to grasp the idea of comparative advantage. The aim of this chapter is to ensure that you fully understand the basis of the theory of trade.

Adam Smith (1723-1790) provided the basic building blocks for the construction of the classical theory of international trade. He enunciated the theory in terms of what is called Absolute Advantage model. Another well-known classicist, David Ricardo (1722-1823) articulated it and expanded it further into what is called Comparative Advantage model. The models of Smith and Ricardo together constitute what is sometimes referred to as the Supply Version of the Classical Theory of Trade, because Smith and Ricardo paid almost exclusive attention to considerations of supply or production costs in the determination of terms of trade and the gains from trade. The modern version of the classical theory of trade, however, treats supply and demand with equal weight. John Stuart Mill (1806-1873), another renowned classical economist, was the first to indicate that demand considerations must be incorporated into the Comparative Advantage model. But Mill was not very clear or articulate. Both Marshall and Edgeworth are credited with originating and developing the theory of offer curves, which is a geometric technique of demonstrating the theory of reciprocal demand. All these contributions of Smith, Ricardo, Mill, Edgeworth and Marshall put together would constitute the modern version of the classical theory of Comparative Advantage, which is the oldest and the most famous model of international trade.

3.1 Absolute Advantage Model of Adam Smith

Adam Smith attacked the mercantilist views on what constituted the wealth of nations, and what contributed to 'nation building' or increasing the wealth and the welfare of nations. Smith was the first economist to show that goods rather than gold (or treasure) were the true measure of the wealth of a nation. He argued that the wealth of a nation would expand most rapidly if the government would abandon mercantilistic controls over foreign trade. Smith also exploded the mercantilistic myth that in international trade one country can gain only at the cost of other countries. He showed how all countries would gain from international trade through international division of labour. In Smith's model of international trade, every one will be better off without making any one worse off; this view contrasts sharply with the mercantilist philosophy that a country can be better off only by making other countries worse off. Smith's model of world trade is one of harmony of interests among countries, where free trade, like honesty, would come out as the best policy for all. Let us now discuss Smith's model with the help of an example,

Imagine for the sake of simplicity that we have a world of only two countries and two commodities. Malaysia and India are two such countries. Rubber and textiles are the two commodities. Assume further that in the production of these two goods in the two countries there are constant returns to scale conditions *i.e.* there are constant marginal opportunity cost conditions in both countries in respect of both the goods. Assume further that the production possibilities are such that both countries can produce both the goods if they wish. Finally assume that both the countries are endowed with "*x*" amount of factors of production such that (a) With *x* factors of production, Malaysia can produce *either* 100 units of rubber *or* 50 units of textiles, or any other mix of rubber and textiles, conditioned by the opportunity cost ratio of 2:1 (This means that if Malaysia wants to produce 1 more unit of textiles it will have to give up the opportunity of producing 2 units of rubber; or alternatively, by giving up the opportunity of producing 1 unit of textiles, Malaysia can produce 2 units of rubber) (b) With *x* factors of production India can produce *either* 50 units of rubber *or* 100 units of textiles; or some other combination of rubber and textiles subject to the opportunity cost ratio of 1:2 (This means that India has to give up producing 1 unit of rubber in order to produce 2 units of textiles; or alternatively, India has to give up the opportunity of producing 2 units of textiles in order to produce 1 more unit of rubber).

Did u know? The vagueness in Mill's principle of what is called as Reciprocal Demand was later removed in the 19th century, first by F.Y. Edgeworth and later by Alfred Marshall.

From the above production possibilities (or supply conditions) it is quite clear that Malaysia has an *absolute* advantage in the production of rubber, and India has the *absolute* advantage in the production of textiles. This means there is symmetrical factor distribution between the two countries so that there is scope for specialization in production and also a scope for establishing mutually beneficial trade between the two countries. Let us see how that happens.

First, in a situation of autarky or no trade between the two countries, each country can produce and consume independent of the other country, a combination of rubber and textiles as shown in the following table :

Table 1 : Production and Consumption Levels with Zero Trade

Countries	Commodities		Total Output or GNP (units)
	Rubber (units)	Textiles (units)	
Malaysia	50	25	75
India	25	50	75
World	75	75	150

Let us now examine the second possibility, *i.e.* when the two countries "open" their economies to foreign trade. Opening up of trade gives the two countries an opportunity to specialize in production. Malaysia would specialize in the production of rubber, because in this line of production it has an absolute advantage over the other country, India. And India will specialize in the production of textiles, because she has absolute advantage over Malaysia in this particular line of production. Note carefully that in this case there is a scope for *complete* specialization in production in both the countries. The effect of opening trade between the two countries is shown in the following table :

Table 2 : Production Levels After International Trade

Countries	Commodities		Total Output or GNP (units)
	Rubber (units)	Textiles (units)	
Malaysia	100	0	100
India	0	100	100
World	100	100	200

After the trade establishment, Malaysia produces only rubber and no textiles. By using all the "*x*" factors of production, Malaysia now produces 100 units of rubber. This is also the level of new GNP in Malaysia after trade. Compare it with Malaysia pre-trade GNP of 75; the GNP increase has been by 25 units in real terms. In the same way, India would specialize in textile production; by using all "*x*" factors of production to produce only textiles, India will be able to produce 100 units of textiles. India's new GNP level is equal to 100 units (of textiles production). Before trade, India's level of GNP was 75 units, which means that India's GNP also rose by 25 units, thanks to international trade.

As a result of trade, you will notice, the GNP in the two countries went up; this means that both countries became richer after trade as compared to before trade. The world GNP also increased from a pre-trade level of 150 to a post-trade level of 200. There has been complete specialization in production after trade. Both countries have become better off in terms of production (GNP) without making any country worse off. This is for production gains from international trade; and the two countries have been richer.

The two countries are of equal size in terms of GNP or the production capacities. These, then, are the levels of GNP and economic welfare in the two countries in the absence of any trade between them, *i.e.* when they are both "closed" economies.

Notes Malaysia produces and consumes 50 units of rubber plus 25 units of textiles *(i.e.* a total real GNP of 75). India produces and consumes 25 units of rubber plus 50 units of textiles *(i.e.* a total real GNP of 75).

What about consumption gains of trade ? After trade, have the consumers in the two countries been happier as a result of their countries becoming richer and more specialized in terms of production? This depends on how the gains from production are distributed between the two countries. In other words, the consumption gains to the two countries depend upon the terms of trade *i.e.* how many units of rubber exchange for one unit of textiles between India and Malaysia.

(a) Suppose the terms of trade are fixed at 1:1, *i.e.* Malaysia and India agree to exchange 1 unit of rubber for 1 unit of textiles. Then, depending upon the taste pattern in the two countries and upon how much or how little they want to trade each other's goods, the consumption gains can be determined. If the two countries want to consume all that they have produced, it means that their consumers have no taste for the product of the other country. Then there will be no trade between them. Nonetheless the two countries will have had production gains (but note, however, that such a condition of production specialization could have been created even without international trade). We describe this situation as one where the consumers have an extreme bias towards the product of their own country, and such situations are unlikely to exist.

Supposing now, that the consumers in both the countries want to consume some mix of both the goods; then Malaysia could export, say, 40 units of rubber in exchange for 40 units of textile imports from India (at 1:1 terms of trade). The resulting situation will be like what it is in Table 3 as given follows :

Table 3 : Consumption Shares After International Trade

Countries	Commodities		Total Consumption (units)
	Rubber (units)	Textiles (units)	
Malaysia	60	40	100
India	40	60	100
World	100	100	200

Malaysia, after trade, has produced 100 units of rubber *(See* Table 2). Consumers in Malaysia wish to consume 60 units of rubber, which means that this country can export 40 units of rubber to India. At international terms of trade of 1:1, Malaysia exports 40 units of rubber in exchange for 40 units of textile imports from India. After trade, the consumers in Malaysia are able to consume a combination of rubber and textiles of 60 and 40 *(See* Table 3). Compare these consumption levels with those of pre-trade in Table I here it was 50 units of rubber and 25 units of textiles. Clearly, consumers have gained in terms of both rubber and textile consumption after trade, as compared to before trade.

Similarly for the other country, India imports 40 units of rubber and exports 40 units of textiles at the same terms of trade. India's post-trade consumption of rubber and textiles is 40 and 60 *(See* Table 3) whereas it was 25 and 50 before trade *(See* Table 1). Therefore, as a result of trade, consumers in this country have also gained.

Note that the consumers in both the countries have gained—and they have gained in equal terms; *e.g.,* Malaysia' consumption gain is 25, which is exactly equal to India's consumption gain also of 25 (compare post-trade consumptions of rubber and textiles in the two countries, in Table 3 with their pre-trade consumptions of the two goods in Table 1). This equal share in consumption gains between India and Malaysia has been made possible by 1:1 terms of trade. The gains from trade (in terms of consumption) depend upon the terms of trade. In this particular case, the 1:1 terms of international trade fall exactly between the two opportunity cost ratios (or internal cost ratios) in the two countries. This is the reason why the two countries' share of consumption gains is absolutely equal.

(b)   Let us now assume that the international terms of trade are 2:1 *(i.e.* two units of rubber are exchanged for one unit of textiles in the international market). This would result in a situation of consumption gains such as the one represented in the following Table :

Table 4 : Consumption Shares After International Trade

Countries	Commodities		Total Consumption (units)
	Rubber (units)	Textiles (units)	
Malaysia	50	25	75
India	50	75	125
World	100	100	200

After trade, Malaysia produces 100 units of rubber; it retains 50 units of it for its own consumption and exports the other 50 units to India. At the new terms of trade of 2:1, Malaysia exports 50

units of rubber and imports 25 units of textiles. This means, Malaysia's consumption share of rubber and textiles is 50 + 25 and India's consumption share of the two goods would be 50 + 75, as you can see in Table 4 above. Compare the situation in Table 4 with the situation in Table 1 and you will notice that Malaysia's consumption of rubber and textiles, before or after trade, has remained the same. This is in spite of production gain of 25 units in Malaysia. India's consumption share, however, after trade has gone up to 50 + 75 as compared to 25 + 50 before trade, *i.e.* India's consumption gain has been equal to 50 *(i.e.* 50 + 75 after trade *minus* 25 + 50 before trade). The entire world production gain has been 50 *(i.e.* 200 after trade *minus* 150 before trade), and all of that world production gain has gone entirely to India. Malaysia's consumption gain has been zero.

This means that the terms of trade *(viz.* 2:1 terms) have been in favour of India alone. They are extremely unfavourable to Malaysia. How ? Note here, that the terms of trade at 2:1 are exactly equal to the internal cost ratio in Malaysia *(i.e.* 2:1). What it means is that, for Malaysia, the cost of importing one unit of textiles (which is 2 units of rubber) is exactly the same as the cost of producing one unit of textiles domestically (which is 2 units of rubber). Therefore the rule governing the consumption gain from trade is as follows : If the international terms of trade are equal (or closer) to the internal cost ratio of any country, then that country will not gain at all (or gain less) from international trade regardless of whether that country has achieved production gains from trade. Let us now take another situation. The reverse of what we have taken here, and show the results.

(c)   Let us now assume that the international terms of trade are 1:2 *(i.e.* one unit of rubber sells for 2 units of textiles in the international market), and these terms of trade are equal to the internal cost ratio in India *(viz.* 1:2). In this situation all the consumption gains must go to the other country—Malaysia, reducing India's consumption gain to zero. Consumers in India would, then, consume the same combination of rubber and textiles after trade as they did before trade. This is despite India's achieving production gains of 25 after international trade as shown in Table 5 below :

Table 5 : Consumption Shares after International Trade

Countries	Commodities		Total Consumption (units)
	Rubber (units)	Textiles (units)	
Malaysia	75	50	125
India	25	50	75
World	100	100	200

After trade, India produces 100 units of textiles—a production gain of 25. It retains 50 units for its own consumption and exports the remaining 50 units of textiles to Malaysia. At the present international terms of trade of 1:2, India will receive 25 units of rubber imports for an export of 50 units of textiles to Malaysia. This means that after trade, India's consumption shares are 25 + 50 of rubber and textiles, and Malaysia's shares are 75 + 50 of the two goods. Compare these figures (in Table 5) with those of pre-trade (in Table 1) and you will notice that India's consumption gain has been zero; all the consumption gains have gone to Malaysia. This is because of the terms of trade (which are 1:2) being exactly equal to the internal cost ratios in India. That means, for India, the cost of importing the good has been the same as the cost of producing that good domestically.

Thus, the gains from trade depend upon the terms of trade. As the terms of trade approach nearer to the domestic cost ratios (or internal terms of trade), the country concerned will start experiencing adverse trading conditions internationally and begins to gain less and less from international trade. Internal cost ratios in the two countries, in a model of world trade where there are only two countries, fix the upper and the lower limits to the international terms of trade. As long as the terms of trade lie

somewhere between the two internal cost ratios (of the two countries) both the countries will share the gains from trade—equally or unequally, depending upon whether the terms of trade are exactly between the two cost ratios or whether they are closer to one country's internal cost ratio or the other country's internal cost ratio. If the terms of trade are closer to the internal cost ratios in Malaysia then Malaysia will gain less than the other country would; if reverse is the case, then India will gain less than Malaysia.

It is important to emphasize that production gains alone are not sufficient to determine the profitability of international trade from the standpoint of an individual member country's welfare gain from trade. Production gain is the GNP gain or income gain. How the consumption gains are determined is crucial in determining whether the economic well-being (or the standard of living measured by consumption gains) of a member country has gone up as a result of international trade. International terms of trade therefore, play a very important part in determining the welfare gains from trade. International trade would be beneficial and profitable for a country only if it results in consumption gains. Production gains alone do not constitute profitable trade from the standpoint of the country concerned.

Today we often hear the complaint that the international terms of trade are very unfavourable to the LDCs. We hear that the advanced industrial countries have been receiving extremely favourable terms of trade. If this is true then the benefits of international trade are going largely to the rich countries. This would aggravate global income inequalities which are already astounding. In that case, international trade would be acting as a source of international inequalities of income and wealth as between the developed and the developing countries of the world. International trade, no doubt, creates prosperity through interdependence. But interdependence can work both ways. We must be careful with the nature of interdependence. The producer and the consumer and interdependent, but the bargaining strength of the consumer may be weaker. What the consumer would be seeking, in fact, is interdependence with equity. That characterizes what most developing countries are concerned about—a greater balance, a greater equity in the relationship between developing and the developed countries. Interdependence, resulting from international division of labour, does not necessarily guarantee equitable trade relationship among nations. Free trade, based on perfect competition, may result in unequal distribution of gains among the countries of the world, even while it increases world prosperity in a spectacular way. It is, therefore, very important that the international terms of trade are just and equitable because profitability of trade is dependent on terms of trade.

Adam Smith's model of classical theory, like the rest of the classical theory of trade, emphasizes the gains from trade *i.e.*, the classical theory is a contribution to welfare economics; and welfare economics is value-loaded. In the classical trade theory the welfare of every individual unequivocally improves with trade; and in the limiting case where a large country trades with a small country, the small country gains more from trade than does the large country. Because, the equilibrium terms of trade nearly coincide with the large country's pre-trade internal price ratio. Thus, in the classical theory the introduction of trade does not make anybody or any country worse off. Not only do all countries gain from trade, in the classical model, but also small countries gain more than the large countries do, emphasizing greater equity inherent in international trade mechanism. Unlike mercantilists who saw conflict of interest among nations, the classical economists saw only harmony of interest among trading nations. This is what led them to pronounce that some trade is better than no trade, more trade is better than less trade, and free trade is better than restricted trade. The economists of the Third World today, however, are not so sure about this. In their view, international trade discriminates against countries which are poor, less developed or too small to exercise any bargaining power with those who are rich, powerful and well developed. This is part of the current theme of what is called as "North-South Dialogue" or the "New International Economic Order".

Getting back to the classical trade theory, it is enough to note that Adam Smith showed convincingly how countries could gain from trade. In his model, to recapitulate, one country has an absolute advantage over the other country in one line of production, and the other country has an absolute advantage over the first country in the other line of production. The two countries' internal pre-trade cost ratios are not the same. International terms of trade would lie somewhere between the opportunity

cost ratios of the two countries, so that both the countries gain from trade in terms of production (or GNP) as well as consumption (or economic welfare) Government non-intervention is a condition, and free trade must be guaranteed.

3.2 Comparative Advantage Model of David Ricardo

Ricardo went even further, and he argued that even if the countries did not have absolute advantage in any line of production over the others, international trade would be beneficial, bringing gains from trade to all the participating countries. Ricardian model is termed as *comparative advantage* model, as opposed to Smith's model of *absolute advantage.* Ricardo's model is a further refinement of Smith's model. Let us now discuss Ricardo's model.

Task   What is absolute advantage?

Once again, let us assume a world of two countries and two commodities. Malaysia and India are the two countries, rubber and textiles are the two commodities. The production possibilities in the two countries are such that both countries can produce both the goods if they wanted; this means that dependence on each other is not inevitable, because the two countries can produce and consume some combination of the two goods, working in isolation (closed economy). Thus far Ricardian model is similar to Smith's model, but the differences arise from here on. In the Ricardian model we assume that one country has the absolute advantage over the other country in *both* the lines of production, and the other country has the absolute disadvantage in *both* the lines of production (contrast this with Smith's model, where one country has absolute advantage in one line, and the other country in the other line). This is in terms of absolute advantage. In terms of relative or comparative advantage, Ricardo assumes that the first country (which has absolute advantage in either line of production) has a greater comparative advantage in one line compared with the other line, in which its comparative advantage is smaller; and the other country's *(i.e.,* the one which has no absolute advantage in either line of production) comparative disadvantage is smaller in the second line compared with the first line of production, where its comparative disadvantage is greater. In brief, one country's comparative advantage is greater in one line of production, and the other country's comparative disadvantage is smaller in the other line of production. International trade would bring production and consumption gains, when these two countries enter into trade with each other. Let us see, with the help of a numerical model, how that happens.

The following table shows the production possibilities in the two countries.

Table 6 : Production Possibilities in India and Malaysia

Countries	Commodities		International Opportunity (Cost Ratios)
	Textiles (units)	Rubber (units)	
India	120 *or*	120	1:1
Malaysia	40 *or*	80	1:2

With "*x*" factors of production, India can produce 120 units of textiles or 120 units of rubber, or any combination of textiles and rubber at the constant opportunity cost ratio of 1:1 *i.e.* India can produce 1 unit of rubber (or textiles) by giving up the opportunity of producing 1 unit of textiles (or rubber). India is equally efficient in the production of the two commodities.

Malaysia, on the other hand, is equally inefficient in either line of production compared with India; because, with *"x"* factors of production Malaysia can produce either 40 units of textiles (compared with India's 120 units) or 80 units of rubber (compared with India's 20 units), or any combination of

textiles and rubber at the constant opportunity cost ratio of 1:2 *i.e.* Malaysia has to give up the opportunity of producing 2 units of rubber in order to produce 1 unit of textiles, or alternatively it has to give up half a unit of textiles, in order to produce 1 unit of rubber. Notice here that the internal cost ratios in the two countries are not the same. To produce 1 unit of rubber or textiles, India has to give up 1 unit of the other alternative commodity foregone. The cost of producing a unit of either commodity is the *same* in India. In the case of Malaysia, however, the cost of producing the two commodities is *not the same.* Because, to produce 1 unit of rubber, Malaysia has to give up half a unit of textiles; but to produce 1 unit of textiles, it is necessary to give up 2 units of rubber. This means that the unit cost of producing rubber is less than the unit cost of producing textiles, when we measure unit costs in terms of the units of alternative commodity foregone. And as long as the internal cost ratios in the two countries are different, there is scope or potential of gains from international trade between the two countries.

In Table 6 you will notice that India has absolute advantage over Malaysia in the production of both the goods, and Malaysia has absolute disadvantage in respect of both the goods. This is as far as absolute advantage and disadvantage is concerned. In terms of relative or comparative advantages and disadvantages, we have the following things to say, *viz,* (a) India's comparative advantage over Malaysia is greater in the production of textiles (3:1) as compared to rubber (1.5:1). Therefore, India should specialize in the production of textiles rather than rubber, although India can produce both the goods equally efficiently (at a cost ratio of 1:1); (b) Malaysia's comparative disadvantage, in relation to India, is lower in the production of rubber (1:1.5) as against textile (1:3). In addition, Malaysia can produce rubber at a far lower cost of production than textiles (at a cost comparisons of 1 unit of rubber = ½ unit of textiles and 1 unit of textiles = 2 units of rubber). Hence, Malaysia should specialize in the production of rubber, not because it has absolute advantage over India in this line but because its comparative disadvantage is less in this line of production than in the other line of production *(viz,* textiles).

The theory of comparative advantage suggests that a country should specialize in the production and export of those goods in which *either* its comparative advantage is greater *or* its comparative disadvantage is less : and it should import those goods, in the production of which its comparative advantage is less *or* comparative disadvantage is greater. Thereby, a country would be able to maximize its production (GNP) and its consumption (economic welfare).

We have already indicated that India should specialize in the production and export of textiles, because her comparative advantage is greater in that line of production, and that Malaysia should specialize in the production and export of rubber; because Malaysia's comparative disadvantage is smaller in rubber production. Before we examine the gains from trade for the two countries arising out of such specialization, let us consider what would the GNP and welfare levels be, for the two countries, in the absence of trade. Table 7 below represents this equilibrium under conditions of autarky.

Table 7 : Production and Consumption under Autarky

| Countries | Commodities | | Total Production |
	Textiles (units)	Rubber (units)	and Consumption (units)
India	80	40	120
Malaysia	20	40	60
World	100	80	180

India produces and consumes 80 units of textiles plus 40 units of rubber, for a total real GNP of 120 units. Malaysia produces and consumes 20 units of textiles plus 40 units of rubber, for a total real GNP of 60 units. The World GNP is 180 units (*i.e.* the GNP of India plus the GNP of Malaysia). Table 7, therefore, represents pre-trade equilibrium situation.

If, however, the two countries decide to enter into trade breaking their isolation, there would be International specialization in production, leading to increase in world GNP. In Table 7 below, where

the production gains are shown, note carefully that while there has been gain in terms of World GNP increase, this had come about entirely through production gain in Malaysia, *i.e.* there has been no production gain in or for India.

World GNP has gone up from 180 to 200 after the introduction of trade. This is entirely due to production gains resulting from specialization in Malaysia, after trade. There are no production gains to be derived from specialization as far as India is concerned, because India's level of production or GNP is the same *(viz.* 120 units) both before and after trade. This suggests that small countries tend to benefit more than the large countries from the standpoint of specialization in production resulting from the establishment of international trade. As far as large countries are concerned—India, in this case—they can attain production specialization even without the help of international trade due, mainly, to the large size of their domestic markets. Small countries, on the other hand, need foreign markets in order to achieve specialization in production. Viewed in this light, small countries stand to gain more from trade and specialization than the large countries.

Table 8 : Production Levels after International Trade

Countries	Commodities		World Production or GNP (units)
	Textiles (units)	Rubber (units)	
India	120	0	120
Malaysia	0	80	80
World	120	80	200

Unless both countries stand to gain from trade, there can be no trade between them. Production gains have gone to only Malaysia, and India has no production or GNP gains from trade. This means that India, as well as Malaysia, must have some consumption gains in order that there is mutually beneficial trade between the two countries. The consumption gains for the two countries depend upon how the production gain is shared or distributed between the two countries. In other words, how much (or how little) each country gains from trade, in terms of consumption or welfare, depends entirely on the terms of trade. The role of terms of trade in distributing trading gains can hardly be overstayed. Let us inspect some possibilities below :

1. If the international terms of trade between India and Malaysia are, say, 3:4 *(i.e.* 3 units of textiles have to be exported in order to import 4 units of rubber, or vice versa) then both the countries will share the benefits equally. This is because these international terms of trade *(viz.* 3:4) lie exactly between the two internal opportunity cost ratios of India and Malaysia. The consumption gains resulting from such International terms of trade for the two countries are shown in Table 9 below :

Table 9 : Consumption Levels After International Trade

Countries	Commodities		Total Consumption (units)	Gains in Consumption (units)
	Textiles (units)	Rubber (units)		
India	90	40	130	10
Malaysia	30	40	70	10
World	120	80	200	20

India, after trade produces 120 units of textiles; she consumes 90 units of it by herself and exports the remaining 30 units to Malaysia. By exporting 30 units of textiles, India receives 40 units of rubber as imports from Malaysia at the terms of trade of 3:4. This means, when the

trade transacting is completed, Malaysia will have 30 units of textiles plus 40 units of rubber for its own consumption. Compare India's post-trade consumption of the two commodities in Table 9 (*viz.* 90 + 40) with their pre-trade levels in Table 7 (*viz.* 80 + 40), and you will notice a net consumption gain of 10 units (in terms of its own product, textiles. This is shown in the last column in Table 9. Similarly, compare also Malaysia's post-trade consumption of the two commodities in Table 9 (*viz.* 30 + 40) with their pre-trade levels in Table 7 (*viz.* 20 + 40) you will once again see that Malaysia's net trading gain in terms of consumption has been equal to that of India *i.e.* 10 units (also in terms of India's product, textiles). Both countries have gained equally in terms of consumption (*viz.* by 10 units each), and this is shown in the last column in Table 9.

In this particular case, note two things—(a) World production gain of 20 has resulted entirely from the production gain in Malaysia, and the production gain has been in terms of Malaysia's product, *viz.* rubber; (b) consumption gain for the two countries, 10 each for India and Malaysia, has been in terms of India's product, *viz.* textiles. What is important, however, is that although production gains have gone up only for Malaysia, the consumption gains are distributed equally between Malaysia and India. This is due to equitable terms of trade which are at a point exactly mid-way between the internal cost ratios in India and Malaysia. Note also that there has been complete specialization in both the countries : India produces only textiles using all the available economic resources in the country, and Malaysia produces all the rubber it can, by using all of its resources for rubber production alone. There is full employment in both countries, since all economic resources are fully utilized in the production of goods, International specialization has been extreme and complete.

2. If the international terms of trade are 1:2 (*i.e.* where one unit of textiles will be exported in exchange for 2 units of rubber on the international market), then all the consumption gains would go to India; because these terms of trade are equal to the internal cost ratio in Malaysia. In this situation, therefore, the cost of importing one unit of textiles would be the same as the cost of producing it domestically in Malaysia. As a result, Malaysia would gain nothing from the standpoint of consumption, even though the production gains have taken place only in Malaysia. Table 10 illustrates this possibility below :

Table 10 : Consumption Levels after International Trade

Countries	Commodities		Total Consumption (units)	Gains in Consumption (units)
	Textiles (units)	Rubber (units)		
India	100	40	140	20
Malaysia	20	40	60	0
World	120	80	200	20

India produces 120 units of textiles, keeps 100 units for its own home consumption and exports 20 units to Malaysia. At international terms of trade of 1:2, India would import 40 units of rubber in exchange for 20 units of textile exports to Malaysia. Thus, India's post-trade consumption combination of textiles and rubber would be 100 plus 40, and that of Malaysia 20 plus 40 (as in Table 10). Compare these figures with the pre-trade consumption levels in the two countries (in Table 7) and you will see that India's consumption gain is equal to 20 and that of Malaysia equal to zero, after the establishment of trade between them. All the production gains (*i.e.* of 20 units) arising out of specialization in Malaysia have gone over to India in the form of consumption gains (of 20 units). Malaysia has, so to speak, all the production gains, and India gets all the consumption gains. In Malaysia, as in India, production has been more specialized after trade; and in fact there has been extreme or complete specialization in Malaysia (as well as India) But in spite of production specialization, Malaysia gains nothing from trade, measured in terms of consumption or living standards. Malaysia's failure to get a share in the consumption gains from trade results from terms of trade which are extremely unfavourable to Malaysia. This, once again, underlines the

importance of terms of trade can nullify the advantages to be obtained from international specialization resulting from trade. Strictly from the country's point of view (or national as opposed to global standpoint) it is the consumption gains that matter; and consumption gains for a country are determined solely by international terms of trade. Production gains could, if terms of trade go against the country, go away to the other side *i.e.* to the foreign countries. The Third World countries, which allegedly suffer adverse terms of trade today, are increasingly getting frustrated with the mechanism of international trade which is benefiting the advanced industrial countries disproportionately. A mutually beneficial trade requires, as a pre-condition, a mechanism of international trade (or interdependence) consistent with equity or justice in trading terms.

3.3 Opportunity Cost and the Pure Theory of Trade

So far in this unit, the free trade doctrine has been discussed in terms of alabour theory of value in which the value of a commodity is determined by the amount of labour time used in its production. Following on from Smith and Ricardo, economists in the nineteenth century subsequently modified and finally abandoned the labour theory of value. It was replaced with the familiar economics 'toolbox' of the present day, in which the value of a commodity is related to its market price, which depends not only on supply and cost conditions, but also on demand.

Neo-classical trade theory

The economists who later overturned the labour theory of value were from continental Europe as well as from Britain. Jean Baptiste Say (1767–1832) was French. Though afirm disciple of Smith, he was the first economist to break away entirely from the labour theory of value. He is generally credited as developing the forerunner of formal equilibrium analysis. Of the three 'founders' of the marginal utility school in the late nineteenth century, Jevons was from England, Menger from Austria (Vienna) and Walras from Switzerland (Lausanne).

The 'neo-classical' thinkers, led by Jevons, Menger and Walras, developed theories of an economic system based on large numbers of producers and consumers. Given a competitive market economy, prices would guide consumers and bring about the most efficient allocation of resources in order to maximise society's income. Neo-classical economists also made great use of mathematical and geometric exposition in order to show functional relationships between important variables such as price and quantity demanded. The use of mathematics ensured greater rigour in the development of their theories.

This is the context in which economists have developed the pure theory of trade. The pure theory of trade treats international trade within the framework of neoclassical theory. It carries through to the present day Adam Smith's belief in the invisible hand of the market, competition and the benefits of laissez-faire policy in relation to international exchange. The pure theory abandons the labour theory of value. Instead it is based on rigorous analysis of consumer and producer behaviour.

The pure theory of trade can be developed through a system of equations and this is the most exact way of presenting it. In this unit, however, we rely on a simple geometric exposition instead of on equations.

Opportunity cost

The doctrine of free trade holds good even if we discard the labour theory of value. The Austrian economist Gottfried Haberler first demonstrated this in the 1930s, utilising the concept of 'opportunity cost'. If the concept of the 'indifference curve' is also introduced into the analysis, it becomes possible for the first time to demonstrate the gains in real income from trade. What follows here is a simplified form of the pure theory of trade based on Haberler's *Theory of International Trade* (1933).

Assume two countries, the US and UK, and two commodities, wheat and cloth. The purpose of the analysis is to demonstrate that the UK gains from specialising in the production of cloth in which it has a comparative advantage, and exporting it to the US in exchange for wheat in which it has a comparative disadvantage. The gains from trade come about because the domestic opportunity cost of cloth in terms of wheat differs from the international opportunity cost of cloth and wheat.

Figure 1 shows the UK (country A). The axis O*y* represents units of wheat. The axis O*x* represents units of cloth. If all resources available in the UK are devoted to producing cloth, O*n'* units of cloth will be produced. If all the resources available in the UK are devoted to producing wheat, O*n* units of wheat will be produced. Any point on the curve *nn'* represents a combination of wheat and cloth production, *nn'* is the production possibility frontier for country A. Assuming all resources are fully employed, country A will be producing at some point on the production possibility curve where both wheat and cloth are produced.

Where on the production possibility curve will country A be located ? To answer this question we need information

- on the *preferences of consumers* in country A for wheat relative to cloth, and
- on the *relative prices* of wheat and cloth. Remember, at this stage we have not introduced the possibility of foreign trade.

Information on relative prices is therefore represented by the *domestic price schedule.* Information on preferences is represented by the *community's indifference curve.*

The indifference curve *ii'* in Figure 3.1 shows the two goods, wheat and cloth, and the combinations of wheat and cloth that are equally acceptable to consumers in country A. The price schedule *pp'* shows the relative prices of wheat and cloth, the rate at which they can be traded one for another in country A. In effect, the slope of *pp'* is the domestic opportunity cost.

The 'no foreign trade' or 'autarky' equilibrium is at *e.* Here the marginal rate of transformation in production (the slope of *nn'*) is equal to the marginal rate of substitution in consumption (the slope of *ii'*) and is equal to the domestic opportunity cost (the slope of *pp'*). At *e,* country A produces O*w* of wheat, and O*c* of cloth. This equilibrium represents the most efficient use of resources for both producers and consumers and yields the maximum level of real income in country A.
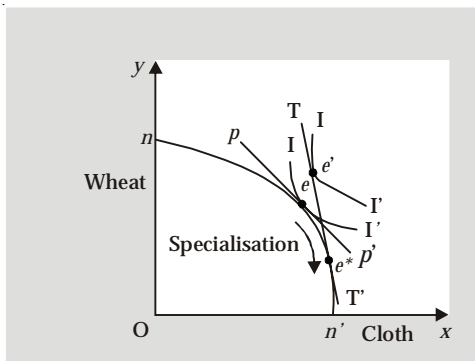
Figure 3.1 : The gains from trade

Gains from trade

We now open up country A to foreign trade. To simplify matters, the analysis uses a partial equilibrium approach, showing the effects of trade on country A only. If we were to introduce country B, as in a general equilibrium approach, more sophisticated geometric tools would be needed.

We know that foreign trade is beneficial if the domestic opportunity cost is different from the international opportunity cost of wheat and cloth. A line TT' is constructed to represent the international terms of trade, *i.e.* the rate at which wheat trades (exchanges) for cloth in the international economy. Because it differs from the domestic opportunity cost it is constructed with a different slope to *pp'.* TT' also indicates how much cloth country A will produce when it is opened up to trade. It must be tangential to *nn'* because after trade, country A must still be somewhere on its production possibility curve, producing both wheat and cloth. The location is indicated by the new post-trade

equilibrium e^*. Country A has moved along the production possibility curve to e^*, where it is producing more cloth and less wheat. It has specialised in cloth at the expense of wheat because it is assumed to have a comparative advantage in cloth.

Exports of cloth trade at the more favourable international opportunity cost represented by TT'. TT' is determined by supply and demand conditions for wheat and cloth in country A and country B. To arrive at the post-trade equilibrium for country A, move out along TT' until a point of tangency is reached with a higher indifference curve II', at e'. At this point, country A's marginal rate of substitution in consumption (slope of II'*)* is equal to the marginal rate of transformation in production (slope of *nn'*), and is equal to the relative prices of wheat and cloth in international markets (slope of TT'*).* II' represents a higher level of real income for country A. Foreign trade, which has led to specialisation and exchange, results in a higher level of real income at the new post-trade equilibrium e'.

Self-Assessment

1. **Choose the correct options**

    (*i*) Which of the following key principles of economics can best explain the benefits from specialization?

    (*a*) The real-nominal principle.

    (*b*) The principle of voluntary exchange.

    (*c*) The principle of opportunity cost.

    (*d*) The marginal principle

    (*e*) The principle of diminishing returns.

    (*ii*) According to the theory of comparative advantage, specialization and free trade will benefit

    (*a*) only the owner of a monopoly.

    (*b*) all trading parties who specialize in the production of the good in which they have a comparative advantage.

    (*c*) only that trading party that has an absolute advantage in the production of all goods.

    (*d*) only the party which specializes the least.

    (*e*) only the party which specializes the most.

    (*iii*) Along the production possibilities curve, specialization and trade usually leads to a move in which the country in question:

    (*a*) produces more of both goods.

    (*b*) produces more of one good and less of another.

    (*c*) produces and exchanges less of both goods.

    (*d*) produces less of both goods but exchanges more of both goods with another country.

    (*iv*) Which of the following IS NOT a reason for the increase in productivity that comes with specialization?

    (*a*) Repetition. (*b*) Continuity.

    (*c*) Innovation. (*d*) Creative destruction.

    (*v*) Market systems are desirable because

    (*a*) they facilitate exchange and specialization.

    (*b*) people cannot be self-sufficient.

    (*c*) exchange cannot take place without markets.

    (*d*) then everyone pays the same amount in taxes.

    (*e*) producers cannot specialize without markets.

(*vi*) Why do people specialize and trade?

   (*a*) Because each person can consume more than they produce by specializing and trading.

   (*b*) Because most people are not self sufficient.

   (*c*) Because most people are good at producing only one thing.

   (*d*) Because people have to have jobs in order to survive.

   (*e*) To exploit their absolute advantages.

3.4 Summary

- Free trade doctrines have a long and fascinating history in Europe. In 1846 Britain repealed the Corn Laws, an historic event which marked the start of the era of free international trade, and lasted until the great depression of the 1870s. The Corn Laws were the duties on imports of grain, which had been in force in England since the middle of the fifteenth century. Other European countries had similar taxes : France, Sweden, Bavaria, Belgium and Holland.

- The reasoning behind the Corn Laws was as follows. Grain, chiefly wheat, is a staple foodstuff, especially important in the diets of labouring people. But its price varies greatly from year to year, depending on the size and quality of harvests. Duties on imports were levied on a sliding scale in order to stabilise the price of wheat. When the domestic price was high because of a poor harvest, duties were lowered to permit imports. When the domestic price was low because of a bumper harvest, import duties were raised.

- The Corn Laws had another important effect. They benefited agricultural interests at the expense of the newly emerging manufacturing sectors. High prices of grain, maintained through restricting foreign supply, increased the value of land.

- Adam Smith attacked the mercantilist views on what constituted the wealth of nations, and what contributed to 'nation building' or increasing the wealth and the welfare of nations. Smith was the first economist to show that goods rather than gold (or treasure) were the true measure of the wealth of a nation. He argued that the wealth of a nation would expand most rapidly if the government would abandon mercantilistic controls over foreign trade. Smith also exploded the mercantilistic myth that in international trade one country can gain only at the cost of other countries. He showed how all countries would gain from international trade through international division of labour.

- Adam Smith's model of classical theory, like the rest of the classical theory of trade, emphasizes the gains from trade *i.e.*, the classical theory is a contribution to welfare economics; and welfare economics is value-loaded. In the classical trade theory the welfare of every individual unequivocally improves with trade; and in the limiting case where a large country trades with a small country, the small country gains more from trade than does the large country. Because, the equilibrium terms of trade nearly coincide with the large country's pre-trade internal price ratio. Thus, in the classical theory the introduction of trade does not make anybody or any country worse off. Not only do all countries gain from trade, in the classical model, but also small countries gain more than the large countries do, emphasizing greater equity inherent in international trade mechanism.

3.5 Key-Words

1. Absolute advantage   : The ability of a country, individual, company or region to produce a good or service at a lower cost per unit than the cost at which any other entity produces that good or service.

2. Comparative advantage   : In economics, the law of comparative advantage refers to the ability of a party to produce a particular good or service at a lower marginal and opportunity cost over another. Even if one country is more efficient in the production of all goods absolute advantage in all goods) than the other, both countries will still gain by trading with each other, as long as they have different relative efficiencies

3.6 Review Questions

1. Discuss comparative advantage model of Ricardo.

2. What is absolute advantage? Discuss.

3. Write a short note on the opportunity cost and the pure theory of trade.

Answers: Self-Assessment

1. (*i*) (*c*)	(*ii*) (*b*)	(*iii*) (*b*)
(*iv*) (*d*)	(*v*) (*a*)	(*vi*) (*a*)

3.7 Further Readings

1. Debreu, G. (1951) : "The coefficient of resource utilization," *Econometrica,* 19, 273–292.

2. Dominguez, D. (2006) : "Lower bounds and recursive methods for the problem of adjudicating conflicting claims," *mimeo.*

3. Foley, D. (1967) : "Resource allocation and the public sector," *Yale Economic Essays,* 7, 45–98.

4. Kalai, E. (1977) : "Proportional solutions to bargaining situations : interpersonal utility comparisons," *Econometrica,* 45, 1623–1630.

Pavitar Parkash Singh, Lovely Professional University

Unit 4 : Modern Theories of International Trade : Theorem of Factor Price Equalization, H-O Theory, Kravis and Linder Theory of Trade

---

CONTENTS

Objectives

Introduction

4.1 The Factor-Price Equalization Theorem

4.2 The Heckscher-Ohlin Theorem (H-O Theory)

4.3 Kravis and Linder Theory of Trade

4.4 Summary

4.5 Key-Words

4.6 Review Questions

4.7 Further Readings

Objectives

After reading this Unit students will be able to:

• Explain the Factor-Price Equalization Theorem.

• Discuss H-O Theory.

• Understand Kravis and Linder Theory of Trade.

Introduction

Modern theories of International Trade can be understand studying the following theory:

1. Resources and Trade (The Eli Heckscher and Bertil Ohlin Model)

    The Heckscher - Ohlin theory explains why countries trade goods and services with each other, the emphasize being on the difference of resources between two countries. This model shows that the comparative advantage is actually influenced by the interaction between the resources countries have (relative abundance of production factors) and production technology (which influences the relative intensity by which the different production factors are being utilized during the production cycle.

    The model starts with the presumption that country A produces two products: food (X) and textiles (Y). These two kinds of production need two different inputs, territory (T) and labour (L), which are available in limited quantities. In the same time, food production (X) requires more land, so it can be said it is territory intensive and textile (Y) production requires more labour, being in this way labour intensive.

    Beginning with these presumptions, the Heckscher-Ohlin model explains the implications trade between two countries A and B has, if the countries produce the same products: food (X) and textiles (Y).

    The relative resource abundance, factors intensity and trade specialization.

Country Product	Inputs and production without trade	The relative abundance and trade specialization in the product for which there is a factor intensity.

	Labour (L)	Territory (T)	L/T	T/L
A	X	Y	X	-
X	20	95	0.21	4.75
Y	10	5	2.00	0.50
Total	30	100	0.30	3.33
B	X	Y	-	Y
X	3	5	0.60	1.66
Y	10	2	5.00	1.20
Total	13	7	1.85	0.53

A country having a bigger offer in a resource than in another is relative abundant in that resource and tends to produce more products that use that resource. Countries are more efficient in producing goods for which they have a relative abundant resource.

According to the Heckscher-Ohlin theory, trade makes it possible for each country to specialize. Each country exports the product the country is most suited to produce in exchange for products it is less suited to produce. In our case, country A is relative abundant in territory (T) and will specialize in producing food (X) and country B is relative abundant in labour (L) so it will specialize in producing textiles (Y). In this case, trade may benefit both countries involved.

The changes in relative prices of goods have a powerful effect on the relative income obtained from the different resources. International trade also has an important effect on the distribution of incomes.

2. Specific Factors and Income Distribution (Paul Samuelson - Ronald Jones Model)

There are at least two reasons why trade has an important influence upon the income distribution:

   (*a*) resources can't be transferred immediately and without costs from one industry to another.

   (*b*) industries use different factors and a change in the production mix a country offers will reduce the demand for some of the production factors whereas for others it will increase it.

Paul Samuelson and Ronald Jones, two American economists, elaborated a trade model based on specific factors.

This is a tri-factorial model because it is based on 3 factors: labour (L), capital (K) and territory (T). Products like food (X) are made by using territory (T) and labour (L) while manufactured products (Y) use capital (K) and labour (L). From this simple example it is easy to observe that labour (L) is a mobile factor and it can be used in both sectors of activity, while territory and capital are specific factors.

A country having capital abundance and less land tends to produce more manufactured products than food products, whatever the price, while a country with a territory abundance tends to produce more food. If the other elements are constant, an increase in capital will mean an increase in marginal productivity from the manufactured sector, while a rise in the offer of territory will increase the production of food in the detriment of manufacturers.

When the two countries decide to trade, they create an integrated global economy whose manufacture and food production is equal with the sum of the two countries' productions. If a country doesn't trade, the production for a good equals the consumption.

The gains from trade are bigger in the export sector of every country and smaller in the sector competed by imports.

3. The Standard Model of Trade (Paul Krugman - Maurice Obsfeld Model)

The standard model of trade implies the existence of the relative global supply curve resulting from the production possibilities and the relative global demand curve resulting from the different preferences for a certain good.

The exchange rate (the rapport between the export prices and the import prices) is determined by the crossing/intersection between the two curves, the relative global supply curve and the relative global demand curve.

If the other elements remain constant, the exchange rate improvement for a country implies a substantial rise in the welfare of that country.

4. The Competitive Advantage (Michael Porter's Model)

    The chain value

    Main Activities

    (cost advantage) -Logistics--- production-marketing---services

    Support activities

    quality advantage) ---Firm infrastructure - HR management--stock supply-Technological development

    Competitive Advantage

    Michael Porter identified four stages of development in the evolution of a country:

    - Development based on factors
    - Development based on investments
    - Development based on innovation
    - Development based on prosperity

    The theory is based on a system of determinants, called by the author "diamond":

    - The capacity of internal factors
    - The specific of the domestic market
    - The links between the industries
    - Domestic competition environment

We will discuss each of the two theorems in detail.

4.1 The Factor-Price Equalization Theorem

The factor-price equalization theorem has enjoyed far less limelight than the Heckscher-Ohlin theorem. Nevertheless, it has attracted considerable attention from well-known economists. Heckscher (1919) stated that free trade equalizes factor rewards completely. Ohlin (1933), on the other hand argued that full factor-price equalization cannot occur in practice. Ohlin asserted that free trade brings about only a tendency towards factor-price equalization, and only a partial factor-price equalization is possible. The later models by Stolper and Samuelson (1941) and Uzawa (1959) also support partial equalization thesis. The later works of Samuelson (1948, 1949, 1953) and of Lerner (1953) make out a case for complete factor-price equalization.

In order to demonstrate how the fector-price equalization takes place as a result of international trade, we will use our model of two countries (country A and B), two commodities (goods X and Y) and two factors of production (capital K, and Labour, L). Before trade (*i.e.* in a situation of autarky) we have the following situations : (1) In country A, a labour surplus country, labour is abundant and cheap and capital is scarce and expensive. Therefore, the K/L (or capital labour ratio) is rather low. And once the trade is opened up, labour becomes relatively scarce and the price of labour will go up. Similarly, capital becomes relatively abundant and hence the cost of capital will go down. This follows directly from the Heckscher-Ohlin theorem that the labour surplus country will specialize in the production and exports of labour-intensive goods. In other words, the abundant factor becomes scarce and the scarce factor becomes abundant, relatively so that (after trade) the K/L will go up in country A. (2) In the capital surplus country—country B—the pre-trade situation is such that capital is abundant and cheap, and labour is scarce and therefore expensive. The K/L will be high before trade. But after trade, this country will specialize in the production of capital-intensive goods, so that the demand for capital will rise relative to that of labour. This will result in capital becoming scarce and the cost of

capital going up, and labour becoming abundant and the cost of labour going down. The K/L will drop as a result.

In brief, we start with a low capital-labour ratio in country A and a high capital-labour ratio in country B. This is before trade. But after trade capital ratio will rise in country A and fall in country B until the capital ratios are equalized in the two countries. This is the process by which the factor prices (capital-labour ratios) in the two countries are equalized as a result of trade. Note that this factor-price equalization is brought about without the movement of factors of production between the two countries. What brings about this factor-price equalization, then, is the international trade mechanism. It is in this sense that we can argue that international trade in goods and services, is a substitute for international labour and capital movements (or factor mobility). We shall now explore the meaning and process of how this factor-price equalization will, in fact, be brought about. We will do it with the help of Edgeworth-Bowley box diagrams.

The points of origin for Good X and Good Y are as shown in the diagram. Capital and labour are measured along the horizontal and vertical sides of the above box diagram. Obviously, it is a labour surplus country since it has more supply of labour than of capital.

There are three possibilities with regard to the capital-labour ratio (K/L) in the production of good X and Good Y, and they are as follows :

(a) If the optimum-efficiency locus (or the contract curve) is a linear straight line, such as AB, the capital-labour ratio in the two goods will be equal and remain so regardless of whether more of X is produced or more of Y is produced. In other words, whether we produce at point 1 or 2 or 3 on the line AB, the capital-labour ratio in good X (K/L_x) will be equal to the capital-labour ratio in good Y(K/L_y), which are shown by the equality of the angles of the size of H and G. (H = G).
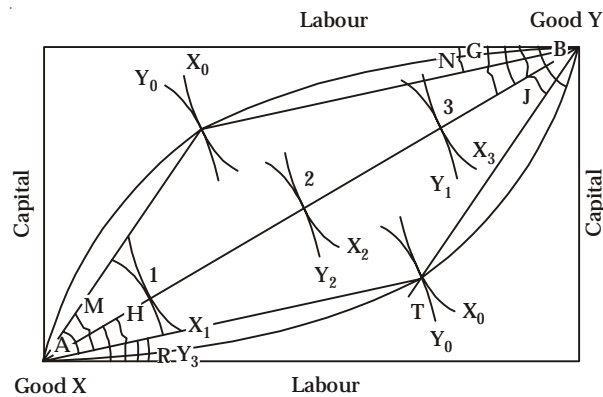
Figure 4.1 : Capital-Labour ratios in Good X and Good Y.

(b) If the contract curve is non-linear of the type represented by the line AFB (in Figure 4.1), Good X will be a capital-intensive good and Good Y will be a labour-intensive good. Because K/L in X (K/L_x represented by the size of the angle M) is higher than the K/L in Y (K/L_y represented by the size of the angle N).

(c) Finally, if the contract curve is non-linearly shaped like the sagging line ATB in the Figure, we find that good X is labour-intensive (low K/L_x, equal to the size of the angle R) and good Y is capital-intensive (higher K/L_y, equal to the size of the angle J).

Throughout what follows, we will assume that the contract curve is of the type represented by the non-linear line ATB so that good X is labour intensive and good Y is capital intensive. Even if the country had more capital and less labour, good X will remain labour-intensive and good Y capital-intensive, so long as we have a contract curve of the shape of ATB as shown in Figure 4.1.

This would mean that in our model (i) country A is a labour-surplus country and country B is capital-surplus country, (ii) good X is labour-intensive and good Y is capital-intensive in both the countries, regardless of the differences in factor proportions and factor prices in the two countries. We shall now turn to the process of factor-price equalization between the two countries, as a result of the opening of trade between them. We will first show what happens in the two countries, using separate graphs, and later on put them together in one composite graph.

Labour-Surplus Country's Case

This is shown (In Figure 4.2)

The box ADBC shows factor supplies in country A, a labour surplus country. The contract curve has the shape of ATB. The production isoquants, X_0 and X_1, are for good X; and $Y_0 Y_1$ are the isoquants for food Y. Before trade, the country produces at point T, where X_0 is tangent to Y_0 at factor price ratio represented by the line $P_0 P_0$. At this point (point T) the capital-labour ratio in good X (K/L_x) is equal to the size of the angle TAC, and the capital-labour ratio in good Y(K/L_y) is equal to the size of the angle TBD. The size of TBD is more than the size of TAC, which shows that good Y is capital-intensive and good X is labour-intensive (or $K/L_y > K/L_x$).
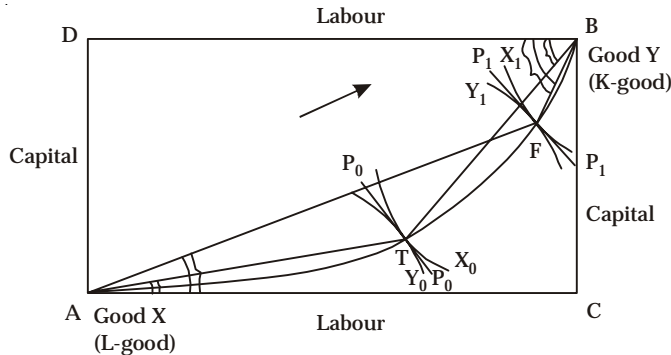
Figure 4.2 : Increase in Capital-Labour Ratio in Labour Surplus Country.

Once trade is opened, this country would specialize in the production (and exports) of good X. This would result in a rightward shift along the contract curve from, say, point T to point F. In moving from point T to F, there is an increase in the production of X and a decrease in the production of Y, which is indicated by an upward movement of the X isoquant and a downward movement of the X isoquant and a downward movement of the Y isoquant. From F we have drawn two rays—FA and FB which would indicate the new capital-labour ratios in the production of X and Y goods at point F. In moving from point T to F, there has been an increase in the capital-labour ratio in X production (from the angle TAC to FAC) as well as Y production (from the angle TBD to FBD). You will, therefore, notice that in country A (a labour surplus country) the capital-labour ratio in both goods production has gone up after the establishment of trade. This is indicated by the change in the slope of factor price line from $P_0 P_0$ to $P_1 P_1$ after trade.

Capital-Surplus Country's Case

The box ABCD in Figure 4.3, represents factor supplies in country B, a capital surplus country.

Country A produced at point T before the introduction of trade. The capital-labour ratio in the production of X was equal to the angle TAC, and in the production of Y it was equal to TBD. Trade results in specialization towards increased production of Y (the capital intensive good), and, therefore, the point of production shifts from point T to, say, point F. As a result of this production-shift, we find that the capital-labour ratios in the production of both X and Y have gone down in this country. In respect of X, the capital-labour ratio has decreased from the angle TAC to FAC; and in respect of Y; it has decreased from TBD to FBD. The change in capital-labour ratios in the production of both X and Y is indicated by the change in the slope of the relative factor-price ratio line from $P_0 P_0$ to $P_1 P_1$.
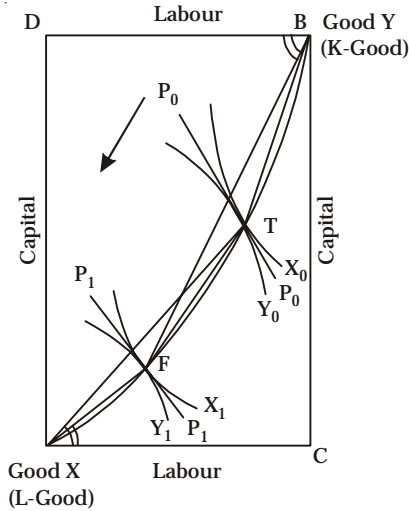
Figure 4.3: Decrease in Capital-Labour Ratio in Capital Surplus Country.

Thus, in the capital surplus country—country A—the capital-labour ratio has decreased in the production of both X and Y goods, as a result of trade. When this is combined with the increase in capital-labour ratio in the production of the two goods in country B—a labour surplus country, we eventually get to a point where the capital-labour ratios (or factor prices) are equalized in the two countries. We will now show this combined result in a composite graph which is drawn in Figure 4.3.

Factor Price Equalization : *A Composite Graph*

In the succeeding graph there are two boxes, one for each country. The box ABCD is for country B, the capital surplus country, and the box AEFG applies to country A which is the labour surplus country. There is a common point of origin at A for good X the labour-intensive good; but for good Y, the capital-intensive good, the point of origin is at point F in case of country A, and at point X in case of country B.

The contract curves, in the two countries, are such that good X is labour-intensive in both the countries and good Y is capital-intensive in both the countries, at all factor price ratios. This applies regardless of the fact that one country is capital rich and the other is labour rich, because the production functions of a given good must be the same throughout in all countries, irrespective what the factor price differences are between the countries.
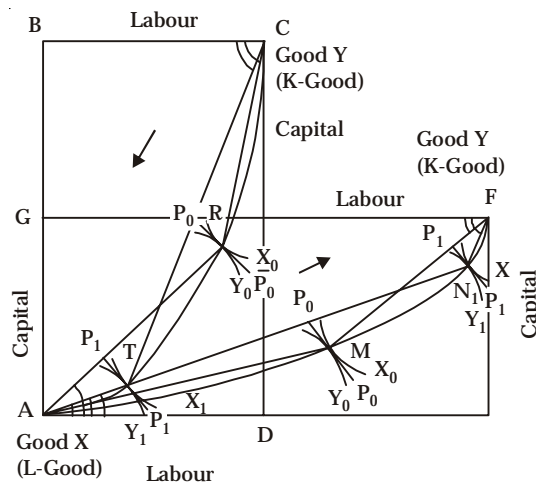
Figure 4.4 : Factor Price Equalization shown jointly.

Under autarky, country A produced at point M and country B at point R. The two countries' factor price ratios were different, as measured by the slope of the line $P_0 P_0$ in the two countries. After the establishment of trade between the two countries, there is a shift in the equilibrium production points : it shifts from point M to N in country A, and from point R to T in country B. $P_1 P_1$ factor price ratio line in country A (at point N) has the same slope as the $P_1 P_1$ line for country B (at point T), which shows that trade has led to equalization of factor prices in country A and B.

You may wish to compare the capital labour ratios in the two goods in the two countries, before and after trade in order to be more clear about how the factor price equalization has resulted after trade.

In Figure 4.4, you will notice the following.

(a)   In country A, at the pre-trade equilibrium production point M, the capital labour ratio is equal to (the angle) MAE in X production, and it is equal to MFG in Y production. After trade, when the country moves to production equilibrium point N, the capital labour ratio increases to (the size of the angle) NAE in X production, and to NFG in Y production.

(b)   In country B, at the pre-trade production point of R, the capital labour ratios in the production of X and Y are measured by the angles RAB and RCD, respectively. The post-trade capital labour ratios in X and Y production (at point T) have gone down to the size (of the angle) of TAE and TCD, respectively.

(c)   After trade, the capital-labour ratio in X production is NAE in country A and TAE in country B. Since NAE = TAE, we can say that the factor-price ratio in X production is equalized between the two countries. In the same way, the capital labour ratio in production, after trade, is TCD in country B and NGF in country A. TCD is equal to NFG, and this shows equalization of factor price ratio in Y production in the two countries.

Thus, we see that international trade brings about equalization of factor prices between countries even in the absence of factor movements between the countries. It is in this sense that we argue that international trade in goods and services, is a substitute for international movement of labour and capital. In other words, international trade brings about equalization of both the product prices and the factor prices. This completes our demonstration of the factor price equalization theorem.

Task   Does factor-price equalization indeed take place in real life ?

Obstacles to Equalization of Factor Prices

Let us remind ourselves that factor price equalization would imply that after the establishment of trade, the capital-labour ratio in the production of a good will be the same in both the countries. This does not mean that the capital-labour ratio in the production of the two goods will also be equalized. This question now is, how far the factor price equalization will conform to the actual reality.

The factor price equalization theorem is based on certain assumptions which are rather unwarranted in real life. The following key points are worth taking note of, because they constitute obstacles to the equalization of factor prices in the real world.

First, the theorem assumes complete free trade, *i.e.* the absence of tariff and non-tariff barriers to trade. It also assumes that there are no transport costs in exporting and importing goods and services between nations. In real world, we know that both exist. Their very existence, therefore, will prevent factor price equalization that free trade could have otherwise brought about. For this reason, Ohlin and others ruled out the possibility of complete equalization of factor prices. They believed that only a partial equalization was possible. Indeed, they argued that there will at best be only a tendency towards factor price equalization.

Complete equalization of factor prices could take place if, and only if, the factors of production were themselves internationally perfectly mobile. In that sense, international trade in goods and services is not a perfect substitute for international factor mobility. It is at best a close substitute.

Secondly, international trade would only lead to partial or incomplete specialization, but by no means to complete specialization in production. This will, therefore, rule out the possibility of complete factor price equalization. At best, we can expect partial equalization of factor prices.

This conclusion is based on the premise that there are diminishing returns to scale conditions in the production of all goods in all countries.

Thirdly, the theorem is based on the assumption of perfect competition and diminishing returns to scale in production. In the real world, however, there is imperfect or monopolistic competition on the one hand, and on the other hand, there are increasing returns to scale in the production of some goods. This would destroy the credibility of the factor price equalization theorem.

Fourthly, for complete factor price equalization to take place the number of factors should not exceed the number of products. In a model of two countries, two factors and two goods, it is possible to show factor price equalization. But, in a real world model of many countries, many factors and many goods and services, it is not possible to argue for a complete equalization of factor prices.

Fifthly, the theorem would collapse once we show that the production functions are not identical in all the countries taking part in international trade. The theorem will not hold good if the factor intensity reversal takes place, because in that event a capital rich country and a labour rich country will export the same good by using different techniques of production suited to their factor endowments. Factor intensity reversals create obstacles to factor price equalization.

Finally, the theorem assumes that factor supplies remain fixed in every country. This is unrealistic, because we do know that the supplies of labour and capital keep constantly changing over time, almost in every country.

We can, therefore, conclude by saying that (a) factor price equalization is not possible in the real world, because the reality of the world does not conform to the assumptions of the factor price equalization model. This does not mean that the model or the theorem is invalid; it only means that the assumptions of the model are unrealistic, which therefore lead the model to draw unrealistic conclusions, (b) given the relative factor immobility between countries, international trade would offer the best chance for equalization of factor prices and factor income between the countries of the world. The factor price differentials would be even greater without international trade than it would be with international trade.

The Stolper-Samuelson Theorem

According to their theorem, the opening of trade (*i.e.* free trade) will benefit the relatively abundant factor and hurt the relatively scarce factor of production. Let us now demonstrate their theorem. Stating their assumptions first.

1. The country, in question, is producing only two goods, say, steel and cloth, with the help of two factors of production, *viz*, capital and labour.

2. Steel is capital intensive and cloth is labour intensive, and the production functions are homogeneous of degree one.
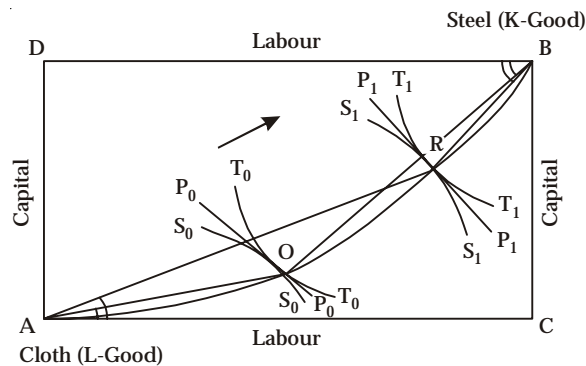
3. The supplies of both factors are fixed.

Figure 4.5 : Stolper-Samuelson Theorem.

4.   The country is assumed to be labour-abundant and capital scarce.

5.   There is perfect competition in both the factor market as well as the commodity market.

6.   International terms of trade remain fixed.

The fixed amount of labour and capital is shown by the Edgeworth box in Figure 4.5.

The origin of cloth isoquant (T_0 and T_1) is A and that of steel (S_0 and S_1) is B. Given the non-linear contract curve sagging below, cloth is labour-intensive and steel is capital-intensive at all production levels of cloth and steel. In the absence of trade, (*i.e.*, autarky) the production takes place at Q on the contract curve, where the cloth and steel isoquants, T_0T_0 and S_0S_0 are tangent to each other. The factor price ratio at Q is represented by the line P_0P_0. At this point, the capital-labour ratio in cloth is equal to the size of the angle QAC, and that of steel equal to the angle QBD.

Did u know?   In 1941, W.F. Stolper and Paul A. Samuelson worked out from the Heckscher-Ohlin theory, a theorem of their own, concerning the effect of trade on income distribution.

When trade opens up, the country will move towards increased production of cloth (the labour-intensive good) and reduced production of steel (the capital-intensive good). The production will now take place at R where the higher cloth isoquant (T_1T_1) is tangential to the lower steel isoquant (S_1S_1), and the new set of factor price ratio at R is indicated by the line P_1P_1. Note that P_1P_1 is steeper than P_0P_0. This indicates an increase in the capital-labour ratios in the production of both cloth and steel. The capital-labour ratios in the production of both cloth and steel. The capital-labour ratio in cloth has gone up from (the size of the angle) QAC to RAC, and in steel, it has gone up from QBD to RBD.

Trade must increase the relative price of cloth which is exported by the country. This is obvious from the following related graph (Figure 4.6). Point Q in Figure 4.6, corresponds to point Q in Figure 4.5. Similarly the line P_0P_0 in the former diagrams corresponds to the line P_0P_0 in the latter diagram. The production equilibrium point R (after trade) in Figure 4.5, has a correspondence with point R in Figure 4.6. You will notice in Figure 4.6, the relative prices of cloth and steel are such that they correspond to the line P_0P_0 which is tangent to the transformation curve at point R. This would mean that the price of cloth has gone up relative to steel.
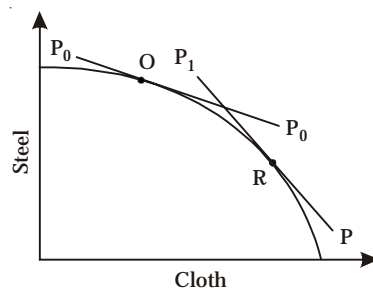
Figure 4.6

As the cloth production expands the resources will be transferred from steel industry to cloth industry. But the steel industry will release such quantities of capital that the cloth industry can only absorb at lower price per unit of capital released. Consequently, the remuneration of capital will have to fall in order that all the released capital is absorbed. On the other hand, the expanding cloth industry will want to employ more labour than the amount of labour released by the contracting steel industry. Consequently, the price of labour is bid higher, and the remuneration of labour will have to go up.

This would, in the end, mean that the absolute income share of labour will go up and that of capital will go down. Trade will, therefore, increase the welfare of the factor of production which is used intensively in the expanding industry, while the factor of production used intensively in the contracting (or import-competing) industry will suffer loss of welfare as a result of trade. In other words, trade increases the welfare of the abundant factor.

4.2 The Heckscher-Ohlin Theorem (H-O Theory)

Recent contributions to the pure theory of international trade have relied heavily on the factor proportions analysis developed by the two Swedish economists, Eli Heckscher (1919) and Bertil Ohlin (1933). According to their theory, the immediate cause of international trade is, the differences in the relative prices of commodities between the countries, and these differences in the commodity prices arise on account of the differences in the factor supplies in the two countries.

The Heckscher-Ohlin model is based upon the following postulates :

1.   There are only two factors of production—labour and capital.

2.   There are only two countries and they differ in factor abundance, *e.g.* one country is capital abundant but labour scarce and the other country is labour abundant but capital scarce. In other words, the two countries differ in factor endowments.

3.   There are only two commodities. Both goods involve the use of both factors. The production functions are such that the relative factor intensities are the same for each good in the two countries. In other words, regardless of what the factor proportions or factor prices are in the two countries, one commodity is always capital intensive in both countries, and the other commodity is labour intensive in both countries.

On the basis of these postulates, the Heckscher-Ohlin theorem predicted that the capital surplus country specializes in the production and exports of capital intensive goods, and the labour surplus country specializes in the production and exports of labour intensive goods. We will now proceed to demonstrate this well-known structure of trade prediction of the Heckscher-Ohlin model.

Factor Abundance Defined : The Two Criteria

In the Heckscher-Ohlin model, the two countries are distinguished by the differences in factor endowments or 'factor abundance' *i.e.* one country is capital abundant (or capital rich) and the other country is labour abundant (or labour rich). The question is, what is meant by 'factor abundance' ? Two alternative definitions have been given for the term 'factor abundance'.

1.   Factor abundance can be defined in terms of factor prices. According to this "price criterion" a country in which capital is relatively cheap and labour is relatively more expensive, is regarded as the capital abundant country, regardless of the physical quantities of capital and labour available in this country compared with the other country. By the same token, a labour abundant country would be defined as one where labour is relatively cheaper and capital is more expensive.

     This criterion takes into account both the supply and demand conditions for the two factors of production in the two countries. Ohlin uses the price criterion of the relative factor abundance, but he argues that the differences in factor prices in the two countries, are due to differences in factor supplies in the two countries. In other words, Ohlin believes that supply factor plays a predominant role in determining the relative factor prices in a country.

2.   Factor abundance can be defined in physical terms. According to the "physical criterion", a country is relatively capital abundant if and only if it is endowed with a higher proportion of capital to labour than the other country. By the same token a labour abundant country is defined as one which has more amount of labour and less amount of capital in physical terms.

This is a pure supply criterion, and it ignores the effects of demand conditions.

These two alternative definitions are not equivalent. The Heckscher-Ohlin prediction with regard to the structure of trade would follow only if we use the price criterion, but it does not necessarily hold good if we use the physical criterion to define factor abundance. Ohlin himself defined relative factor abundance using the price criterion. He thought that if capital is relatively cheap in one country, that

country must be abundant in capital supply; and if labour is relatively cheap in the other country, it must be a reflection of the labour abundance in that country.

It now remains for us to show that one country, say country A, is capital abundant and it exports capital-intensive good, and the other country, say country B, is labour abundant and, therefore, it will export labour-intensive good. We shall examine these Heckscher-Ohlin proposition by using both the definitions of factor abundance—not together but separately.

Price Criterion of Factor Abundance

Starting from the definition of factor abundance in terms of factor prices, it is easy to establish the Heckscher-Ohlin theorem. It is easily demonstrated in the succeeding diagram.

The two factors—capital (K) and labour (L) are measured along the vertical and horizontal axis, respectively. The set of factor price ratios in country A, a capital surplus country, are shown by the parallel lines P_0P_0. The relative steepness of these lines reflects the fact that capital is cheaper and labour is dearer in country A. Similarly, the lines P_1P_1 reflect the factor price ratios in country B. The relative flatness of these lines shows that labour is cheap and capital is expensive in country B, a labour surplus country.

Then we have the two isoquants labelled *aa* and *bb*, and the two isoquants cut each other only once, *i.e.* at point Q. This indicates that there is no reversal of factor intensity, meaning that one commodity is capital intensive in both the countries (K good represented by the isoquant *aa*) and other commodity (*Good* represented by the isoquant *bb*) is labour intensive in both the countries. This is in conformity with the Heckscher-Ohlin assumption that the production functions are identical for each good in the two countries.
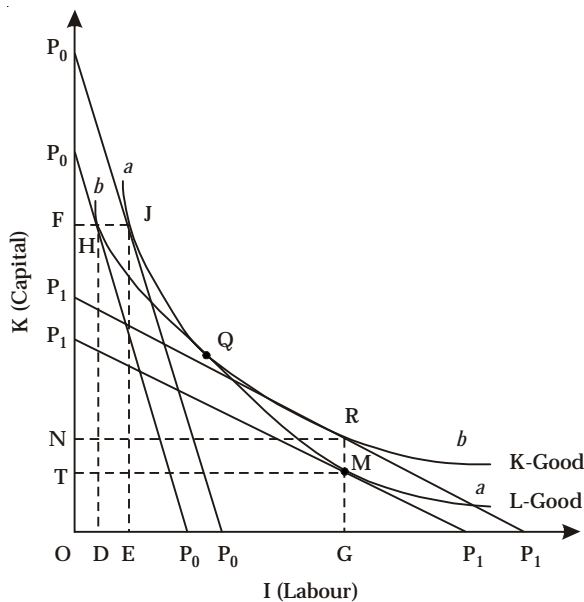
Figure 4.7 : Price criterion of factor abundance.

We can now see how the capital surplus country would export capital-intensive good, and labour surplus country would export labour-intensive good. Take country A first. The cost of producing one unit of K good is made up of HD amount of capital plus HF amount of labour, because at point H there is a tangency between the isocost curve and the isoquant for K good. The cost of producing one unit of L good consists of JE amount of capital (which is equal to HD) but more labour *viz.* FJ amount of labour (as against FH amount needed to produce one unit of K good). In other words, in country A, in order to produce one unit of L good, you need to use the same amount of capital as in K good

(*viz.* HE = JE) but more labour (JF as against HF). This means that country A can produce K good cheaper. Hence the capital surplus country (country A) would specialize in the production and exports of capital-intensive good (K good). This is for country A.

Take the case of country B now. The cost of producing one unit of L good is made up of MG amount of capital plus MT amount of labour, but the cost of producing one unit of K good consists of the same amount of labour *viz.* NR (MT) but more amount of capital *i.e.* RG (as against MG needed to produce one unit of K good). This means that country B can produce L good at a relatively lower cost of production per unit. Therefore, country B (a labour surplus country) would specialize in the production and exports of L good (a labour intensive good).

To sum up : (a) Factor price ratios in country A and B are different, which reflects that country A is capital-abundant and country B is labour-abundant, (b) one commodity is capital intensive in both the countries (*viz.* K good) and the other commodity is labour intensive in both the countries (L good), because point J lies to the right of point H in country A, and in country B, point R lies to the left of point M. (You may draw the vectors from points H, J, R and M to the point of origin, O, and you will see the capital-labour ratios in the production of the two goods in the two countries. This is not done in Figure 4.7 in order not to clutter the graph), and (c) country A can produce K good cheaper, and country B can produce L good cheaper. Therefore, country A exports K good and country B exports L good. The country which is relatively abundant in a given factor of production will export a commodity which involves the use of a relatively abundant factor of production.

Thus, starting from the definition of factor abundance in terms of factor prices, (or price criterion) it is easy to establish the Heckscher-Ohlin theorem. Incidentally, we might also say that reverse of the theorem also holds good, *i.e.* if a country exports capital intensive good, then capital must be its cheaper factor of production. Likewise, if a country exports labour-intensive good, then labour must be a cheaper factor of production in that country.

Physical Criterion of Factor Abundance

Defining factor abundance in physical terms would mean that country A would be capital-abundant and country B would be labour abundant if the following condition holds good :

$$\left(\frac{K_A}{L_A}\right) > \left(\frac{K_B}{L_B}\right)$$

where K_A and L_A are the total amounts of capital and labour, respectively, in country A, and K_B are L_B one the amounts of capital and labour, respectively, in country B.

We will now show that country A, a capital abundant country by the physical criterion of abundance, has a bias in favour of producing the capital-intensive good; and that country B, a labour abundant country will have a production bias in favour of labour-intensive good production. Diagram 8 reflects the nature of these biases in the two countries in respect of the two goods.
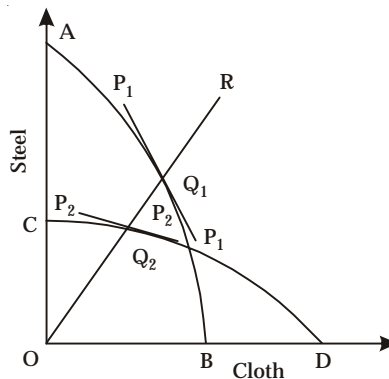

Figure 4.8 : Factor abundance defined in physical terms.

The production possibility curve of country A is AB that of country B is CD. We assume that steel is the capital intensive good and cloth is the labour intensive good. Suppose, the two countries produced the goods in the same proportion—along the ray OR— then country A would produce at Q_1 and country B at Q_2, on their respective production possibility curves. Note that the slope of country A's production-possibility curve at Q_1, is steeper than the corresponding slope of country B at Q_2. Similarly, the commodity price line P_1P_1 is steeper than the line P_2P_2. All this implies that steel is cheaper in country A and cloth is cheaper in country B, if the two countries are producing at Q_1 and Q_2 respectively. Country A would, therefore, tend to expand production of steel and country B would do so for cloth. This means that country A, a capital abundant country, has a production bias in favour of capital-intensive good, steel, while the labour-abundant country, country B, has a bias in favour of producing the labour intensive good, *viz.* cloth.

Does it follow from this that country A would export steel and country B would export cloth ? The answer depends very much upon the demand factors. This gives rise to two possibilities : (1) if the consumption bias and the production bias are towards the same direction, then country A would import rather than export steel, and country B would import rather than export cloth. The Heckscher-Ohlin prediction would then be invalid, (2) if the consumption and production biases are in the opposite direction, then the Heckscher-Ohlin prediction will be valid, *viz.* country A would export steel and country B would export cloth. Let us illustrate these two cases.
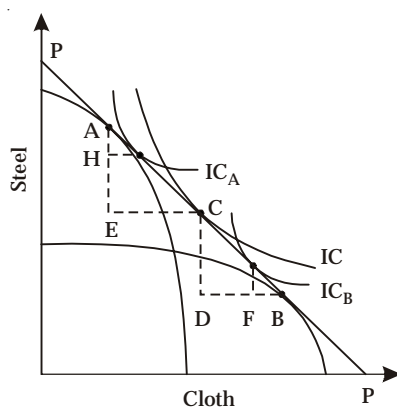
Figure 4.9 : Consumption and production biases in opposite direction.

After the establishment of trade between the two countries, country A's production shifts to point A (towards greater production of steel), and country B's production shifts to point B (towards greater production of cloth). This means that the capital surplus country (country A) specializes in the production of the capital-intensive good (steel), and the labour surplus country (country B) specializes in the production of the labour-intensive commodity (cloth). There is greater degree of specialization but by no means complete specialization, in the two countries, because of the diminishing returns to scale conditions in the two countries in respect of both the goods.

The line PP stands for the international terms of trade line, which is also the relative factor price ratio line after trade is established between the two countries. (Note incidentally that factor prices will be equalized as a result of trade. We shall discuss this later under factor price equalization theorem, which is the second proposition of the Modern theory of international trade).

If, and only if, the demand biases in the two countries are such that we have an indifference curve like IC in Figure 4.9A, the Heckscher-Ohlin theorem will hold good. The two countries produce at points A and B but consume at point C. It is important that consumption point, such as point C, must lie to the right of point A but to the left of point B in order for Heckscher-Ohlin prediction to be valid. In this case, it has happened in Figure 4.9. Country A exports an amount of steel equal to the size AE

and imports cloth equal to the size EC. Country B exports DB amount of cloth and imports CD amount of steel. The capital surplus country, therefore, is exporting capital intensive good and it is importing labour-intensive good.

Similarly, the labour surplus country is exporting labour-intensive good and it is importing capital-intensive good.

In this case, therefore, the Heckscher-Ohlin prediction would be valid. We emphasize that it is important that consumption should take place to the right of where the production is taking place in country A, and to the left of where the production is taking place in country B. Only then the two countries will specialize in the production as well as export of the commodities which involve intensive use of their respectively abundant factors of production. It is by no means necessary that the taste pattern in the two countries must be identical. In Figure 4.9, we have made that assumption in drawing a common indifference curve IC both for country A and country B, but it is not necessary. One can feel free to assume that taste patterns in the two countries are different, if that sounds more realistic. For instance, in Figure 4.9, we have also drawn IC$_A$ and IC$_B$ which represent different demand (or utility) patterns in country A and country B. This would only mean that country A is consuming at point R while it produces at point A, and that country B is producing at point B and consuming at point T. Nevertheless, country A exports steel (equal to AH amount) and imports cloth (equal to HR amount); and country B exports FB amount of cloth and imports TF amount of steel. Therefore, as long as the consumption points lie to the right of where production is taking place in country A and to the left of where the production is taking place in country B, the Heckscher-Ohlin prediction concerning production specialization as well as commodity composition of exports and imports by countries would perfectly hold good.

If, on the other hand, the demand patterns are so unidentical that the indifference curve of country A is tangent at point A (not drawn in Figure 4.9) and the indifference curve of country B is tangent at point B (also not drawn in Figure 4.9), then it would mean that country A and B choose to consume where they produce. There will then be no trade, but a situation of autarky. In such an event, the Heckscher-Ohlin prediction will still be valid but only insofar as it related to production specialization but not structure of trade. There will, in fact, be no trade to speak of.

Now, let us take the other case, *i.e.* when the consumption and production are biased in the same direction. This case is illustrated in Figure 4.10.

Figure 4.10 reproduces the same information as in Figure 4.9, except that in Figure 4.10, the demand in country A is biased toward the capital-intensive good and that in country B the demand is biased toward the labour-intensive good. Therefore, as a result, country A produces at point A, specializing in the production of steel. It consumes at point D, given the utility pattern represented by indifference curve IC$_A$. This means that country A exports EA amount of cloth and imports ED amount of steel. Therefore, country A which is a capital surplus country, is exporting labour-intensive good (cloth) and importing capital-intensive good (steel). This is in direct conflict with the Heckscher-Ohlin prediction concerning the commodity structure of trade.
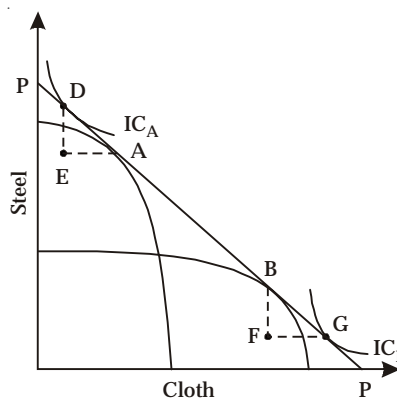

Figure 4.10 : Consumption and production biases in the same direction.

Likewise, country B specializes in the production of cloth; it produces at point G. But it consumes at point G in response to its utility pattern represented by the indifference curve IC_B. Therefore, it exports BF amount of steel and imports FG amount of cloth. Once again we notice that country B, which is a labour-surplus country exports capital-intensive good (steel) and imports labour-intensive good (cloth). The Heckscher-Ohlin prediction is overturned.

In this case, represented in Figure 4.10, we have a situation of what is sometimes described as "demand reversal". Here, not only the two biases—consumption and production—are in the same direction but also the consumption bias more than offsets the production bias. Consumption point D lies to the left of production point A in country A; and in country B the consumption point G lies to the right of the production point B. When such a demand reversal takes place, the capital surplus country would export labour-intensive good and the labour-surplus country would export the capital intensive good. The Heckscher-Ohlin prediction would then be invalidated by the demand reversal.

To sum it up : factor abundance can be defined in two ways in the Heckscher-Ohlin trade model. The two definitions are not equivalent. Only according to the price criterion, the prediction of the model would be valid. If the physical criterion is used, the prediction will be valid only if the demand reversal does not take place.

Critical Evaluation of the Heckscher-Ohlin Theorem

In the area of pure theory of international trade, the Heckscher-Ohlin model occupies a very prestigious position. The very fact that many well-known economists like Leontief, Walters, Minhas and others have tried to test the empirical validity of the Heckscher-Ohlin theorem using econometric models, stands as a testimony of the prestige of the model. Bertil Ohlin was also awarded a Nobel Prize in Economics for his contribution to the pure theory of trade in the year 1978.

Although the factor proportions theorem developed by Heckscher and Ohlin provides a thorough and plausible explanation of international trade as compared with the classical comparative advantage model, yet it is not free from criticism. We have already seen how the Heckscher-Ohlin theorem will turn out to be invalid when the demand reversal takes place. We will, therefore, examine only the first two lines of criticism in what follows.

Notes The Heckscher-Ohlin theorem has been criticized mainly along the following three lines: (a) Factor intensity reversal argument; (b) Leontief Paradox, *i.e.* the results obtained by empirical tests conducted by Leontief and others; and (c) Demand reversal argument.

Factor Intensity Reversal Argument

The Heckscher-Ohlin theorem was based on the assumption that the production functions are different for different goods but they are identical for each good in the two countries. This, in other words, meant that one good is capital intensive (with higher capital-labour ratio) and the other good is labour-intensive (with lower capital-labour ratio); but the same good, which is capital-intensive in one country, must be capital intensive in the other country also, and the labour intensive good remains labour intensive in both the countries. This assumption is guaranteed when the two production isoquants—for the capital-intensive and the labour intensive goods—cut each other only once but not more than once. In Figure 4.7, this is shown to happen at point Q. The demonstration in Figure 4.7 is consistent with the Heckscher-Ohlin assumption of non-reversability of factor intensities.

What will be the effect if the factor-intensity reversal takes place ? If it does take place, then the two isoquants would cut each other more than once and the Heckscher-Ohlin theorem would turn out to be invalid. This case is demonstrated in Figure 4.7.

The two production isoquants for steel and cloth cut each other twice in the succeeding diagram—once at point A and the second time at point B. Now let us see the results. The factor price ratios in country A (capital surplus country) are represented by the parallel lines P_0P_0. P_1P_1 represent the factor price ratios in country B (labour surplus country). This is as in Figure 4.7 also.

In Figure 4.10, note the following factors.

(a) In country A, steel is labour-intensive and cloth is capital-intensive. In order to produce one unit of either steel or cloth, country A has to use the same amount of capital but more labour for steel than for cloth. Cloth has a higher capital labour ratio and steel has a lower capital labour ratio. Therefore, a capital rich country like country A would specialize in the production and exports of the capital intensive good, which is cloth. It would import steel which is a labour intensive good. (b) In country B, cloth is a labour intensive good and steel is a capital intensive good. Because, to produce one unit of cloth, it takes a given amount of labour and smaller amount of capital as compared to steel. Steel takes the same amount of labour but more capital per unit of output. In country B, therefore, steel has a higher capital labour ratio than in cloth. Naturally, country B (which is a labour surplus country) would choose to specialize in the production and exports of the labour-intensive good, *viz.* cloth. Country B, therefore, would export cloth and import steel which is capital intensive.
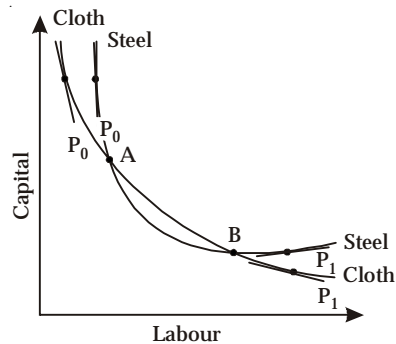
Figure 4.11 : Factor intensity reversal.

In this case of factor intensity reversal, as we saw above, both the countries produce and export the same commodity *i.e.* cloth. In the capital rich country, (country A) it is a capital-intensive product, and in the labour rich country, (country B) it is a labour-intensive product. That means the same product (*viz.* cloth) is capital-intensive in one country but labour-intensive in the other. The same thing applies to steel as well. Steel is a labour intensive product in the capital rich country (country A), and it is a capital-intensive product in the labour rich country (country B). This is a situation of factor intensity reversal. When this takes place, both countries end up producing and exporting the same commodities (cloth) and importing the other commodity steel). This would invalidate the Heckscher-Ohlin prediction regarding the structure of commodity trade.

In Figure 4.11, the two isoquants cut each other more than once, suggesting factor-intensity reversal to the left of point A and to right of point B. For factor intensities to reverse themselves, it is not, however, necessary that the two isoquants intersect each other more than once. Even if they do not cut each other even once, there could be reversal of factor-intensities as shown in Figure 4.12.
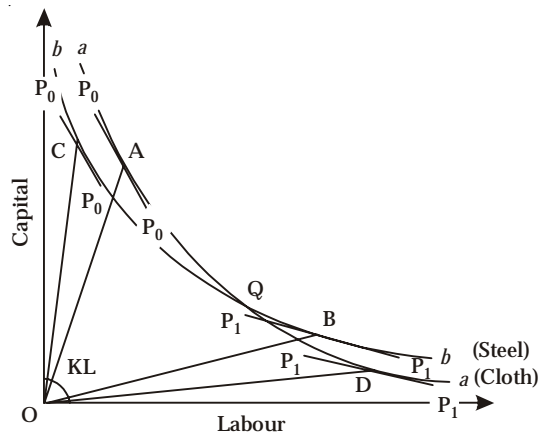
Figure 4.12 : Factor intensity reversal—another case.

In Figure 4.12, the isoquants for steel and cloth, are tangent to each other at point Q; and the factor intensities of the two products reverse themselves on either side of point Q. Given the factor price ratios of country A represented by the lines P_0P_0, cloth has a higher capital labour ratio (K/L) as compared to steel in country A. This is indicated by the vectors OC and OA. In country B, however, steel has a higher K/L ratio than cloth, and this is indicated by the two vectors OB and OD. In other words, cloth is capital-intensive and steel labour-intensive in country A, whereas in country B steel is capital-intensive and cloth labour-intensive. This is another case of factor intensity reversal. Factor-intensity reversal can also be demonstrated as in Figure 4.13.
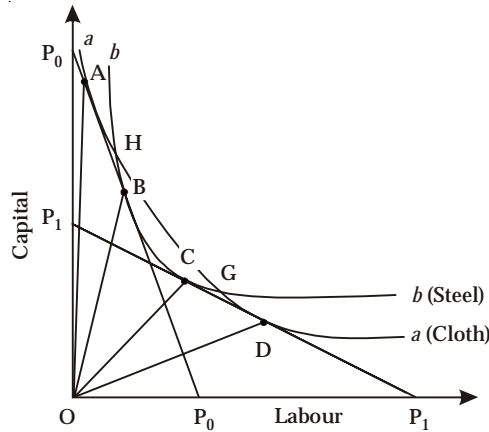
Figure 4.13: Factor intensity reversal once again.

Country A's factor price ratio line is P_0P_0 and country B's similar line is P_1P_1. The isoquant for cloth is the line *aa*, and for steel, it is the *bb* line. The two vectors OA and OB represent K/L ratios in the production of cloth and steel in country A, and the other two vectors OC and OD show the similar ratios for the two goods in country B. It can be easily seen that cloth is capital-intensive in country A but labour intensive in country B, and steel is labour-intensive in country A while it is capital-intensive in country B. The two isoquants cut each other twice—at point H and G.

The assumptions of identical production functions in the two countries and of the non-reversability of factor intensities, are necessary for the validity of the Heckscher-Ohlin prediction. The question, then, is whether factor reversal is common in the real world. Minhas, Leontief, Moroney and others have carried out extensive empirical investigations into this question and their findings are conflicting. Minhas, for example, investigated 24 industries for which comparable data could be obtained for 19 different countries, and found factor reversal in 5 cases. Leontief and Moroney are critical of Minhas's findings, and in fact, Moroney concludes that factor intensity reversal "has much less empirical importance than theoretical interest".

The phenomenon of reversal of factor intensity, if provided widespread, would rob the Heckscher-Ohlin model its predictive significance concerning the structure and direction of commodity trade.

Leontief Paradox and Other Leontief-Type Tests

The first comprehensive and detailed examination of the Hecksher-Ohlin theorem was the one undertaken by Leontief. You will recall that the theory of factor proportions predicted that the capital abundant country exported capital-intensive goods and imported labour-intensive goods, and the labour surplus country did the opposite. It is commonly agreed that the United States is a capital rich and labour scarce country. Therefore, one would expect exports to consist of capital-intensive goods and imports to consist of labour-intensive goods. Leontief made an extensive study of the US structure of trade and the results were startling. Contrary to what the Heckscher-Ohlin theory had predicted,

Leontief's study showed that the US exports consisted of labour-intensive goods and the imports, (or more precisely import competing products) consisted of capital-intensive goods. In Leontief's own words, "America's participation in division of labour in international trade is based on its specialization in labour intensive rather than capital-intensive lines of production. In other words, the country resorts to foreign trade in order to economize its capital and dispose of its surplus labour, rather than vice versa". Leontief's findings are summarized in the following table :

	Exports	Imports Replacements
Capital (US $ in 1947 prices)	2,550,780	3,091,339
Labour (man years)	1,80,313	1,70,004
Capital Labour ratio (US $ per man hour)	13,911	18,185

From the above table, it is obvious that the US exports had a lower capital-labour ratio than the import replacements. Note carefully that these are import replacements produced in the United States as opposed to the actually imported goods in that country. This distinction is important as we will see later.

Leontief's paradoxical results stimulated similar studies for other countries. Two Japanese economists found that Japan's exports embodied more capital and less labour than Japan's import competing goods. Since Japan is a relatively labour-abundant country, this conclusion is inconsistent with the Heckscher-Ohlin prediction. Similarly, an examination of Canada's trade with the United States revealed that Canada's exports were capital intensive and imports labour intensive. This, again, is not in accordance with the Heckscher-Ohlin theory. Furthermore, an investigation of India's trade with the United States discovered exports to be capital-intensive and imports labour-intensive. Considering that India is an extremely labour surplus country, the result is a paradox. Finally, a study of East Germany's trade showed her exports to be capital intensive and imports labour-intensive. East Germany is not really a capital surplus country, this is also a paradoxial result.

4.3 Kravis and Linder Theory of Trade

Non-availability Approach (Kravis Theory of Trade)

The **non-availability** explains international trade by the fact that each country imports the goods that are not available at home. This unavailability may be due to lack of natural resources (oil, gold, etc. : this is *absolute* unavailability) or to the fact that the goods cannot be produced domestically, or could only be produced at prohibitive costs (for technological or other reasons) : this is *relative* unavailability. On the other hand, each country exports the goods that are available at home.

As regards the presence or absence of natural resources this aspect could easily be fitted into the Heckscher-Ohlin model that stresses the differences in relative endowments. A generalized version of the model can be used by adding a factor *natural resources.*

The originality of this approach lies in its second aspect, that is, in the reasons put forward to explain international differences in relative availability. Essentially there are two reasons : *technical progress and product differentiation.*

As regards the first reason, Kravis observes that the stimulus to exports provided by *technological change* is not confined to the reduction costs but also includes the advantages deriving form the possession of completely new products and of the most recent improvements of existing types of goods. In such cases the operation of the demonstration effect of Duesenberry (1949) creates an almost instantaneous demand abroad for the products of the innovating country and thus generates international trade.

As regards *product differentation*, the idea of Kravis is to extend to international trade the results of the theory of monopolistic competition. Different countries produce similar commodities or, more exactly,

commodities that are not substantially different from the point of view of their intended purpose (clothes, automobiles, watches, cameras, cigarettes, liqueurs, etc.). These commodities, however, due to different industrial designs, past excellence, advertising, real or imaginary secondary characteristics and so on and so forth, are considered different by consumers. This creates, on the one hand, a more or less limited monopolistic power of the single producing countries, and on the other a consumers' demand for foreign commodities that they believe different from similar domestic commodities, the result being to create international trade.

Linder Theory of Trade

The **Linder hypothesis** is an economics conjecture about international trade patterns : The more similar the demand structures of countries, the more they will trade with one another. Further, international trade will still occur between two countries having identical preferences and factor endowments (relying on specialization to create a comparative advantage in the production of differentiated goods between the two nations).

Development of the theory

The hypothesis was proposed by economist Staffan Burenstam Linder in 1961 as a possible resolution to the Leontief paradox, which questioned the empirical validity of the Heckscher-Ohlin theory (H-O). H-O predicts that patterns of international trade will be determined by the relative factor-endowments of different nations. Those with relatively high levels of capital in relation to labor would be expected to produce capital-intensive goods while those with an abundance of labor relative to (immobile) capital would be expected to produce labor intensive goods. H-O and other theories of factor-endowment based trade had dominated the field of international economics until Leontief performed a study empirically rejecting H-O. In fact, Leontief found that the United States (then the most capital abundant nation) exported primarily labor-intensive goods. Linder proposed an alternative theory of trade that was consistent with Leontief's findings. The Linder hypothesis presents a demand based theory of trade in contrast to the usual supply based theories involving factor endowments. Linder hypothesized that nations with similar demands would develop similar industries. These nations would then trade with each other in similar, but differentiated goods.

Empirical tests

Examinations of the Linder hypothesis have observed a **"Linder effect"** consistent with the hypothesis. Econometric tests of the hypothesis usually proxy the demand structure in a country from its per capita income : it is convenient to assume that the closer are the income levels per consumer the closer are the consumer preferences. (That is, the proportionate demand for each good becomes more similar, for example following Engel's law on food and non-food spending.)

Self-Assessment

1. **Choose the correct options:**

   (*i*) What is the most important fact about U.S. international trade in the after-war period?

      (*a*) Imports exceed exports.

      (*b*) Both imports and exports grew significantly as a share of GDP.

      (*c*) Exports and imports are a large share of U.S. GDP.

      (*d*) Exports exceed imports.

   (*ii*) Which of the following is NOT true?

      (*a*) Small countries depend more on trade than large countries.

      (*b*) U.S. imports exceed U.S. exports.

      (*c*) Imports cannot exceed exports for an extended period of time.

      (*d*) Economists believe that international trade is beneficial for all countries involved in it, in most cases.

(*iii*) The term "gains from trade" describes:

Producer surplus.

  (*a*) The fact that when two countries trade, both are better off.

  (*b*) Profits made by businessmen involved in international trade.

  (*c*) The income of middlemen in a transaction.

  (*d*) Consumer surplus.

(*iv*) Why do some people argue against free international trade?

  (*a*) Trade alters the distribution of income between broad groups of people.

  (*b*) There is disagreement on whether or not there are gains from trade.

  (*c*) Free trade threatens our country's security.

  (*d*) The U.S. is a large country and therefore does not gain from international trade.

(*v*) Which of the following theories was proposed by David Ricardo?

  (*a*) Theory of differences in factor endowments.

  (*b*) Theory of differences in labor productivity.

  (*c*) Theory of random components determining the pattern of trade.

  (*d*) Theory of differences in climate and resources.

(*vi*) What are most trade policies driven by?

  (*a*) Conflicts of interest within nations.

  (*b*) Conflicts of interest between nations.

  (*c*) Disagreements on the prices of major commodities.

  (*d*) Disagreements regarding who should produce certain products.

4.4 Summary

- Heckscher (1919) stated that free trade equalizes factor rewards completely. Ohlin (1933), on the other hand argued that full factor-price equalization cannot occur in practice. Ohlin asserted that free trade brings about only a tendency towards factor-price equalization, and only a partial factor-price equalization is possible. The later models by Stolper and Samuelson (1941) and Uzawa (1959) also support partial equalization thesis. The later works of Samuelson (1948, 1949, 1953) and of Lerner (1953) make out a case for complete factor-price equalization.

- In order to demonstrate how the fector-price equalization takes place as a result of international trade, we will use our model of two countries (country A and B), two commodities (goods X and Y) and two factors of production (capital K, and Labour, L). Before trade (*i.e.* in a situation of autarky) we have the following situations : (1) In country A, a labour surplus country, labour is abundant and cheap and capital is scarce and expensive. Therefore, the K/L (or capital labour ratio) is rather low. And once the trade is opened up, labour becomes relatively scarce and the price of labour will go up. Similarly, capital becomes relatively abundant and hence the cost of capital will go down.

- Recent contributions to the pure theory of international trade have relied heavily on the factor proportions analysis developed by the two Swedish economists, Eli Heckscher (1919) and Bertil Ohlin (1933). According to their theory, the immediate cause of international trade is, the differences in the relative prices of commodities between the countries, and these differences in the commodity prices arise on account of the differences in the factor supplies in the two countries.

- In the Heckscher-Ohlin model, the two countries are distinguished by the differences in factor endowments or 'factor abundance' *i.e.* one country is capital abundant (or capital rich) and the

other country is labour abundant (or labour rich). The question is, what is meant by 'factor abundance' ? Two alternative definitions have been given for the term 'factor abundance'.

- The first comprehensive and detailed examination of the Hecksher-Ohlin theorem was the one undertaken by Leontief. You will recall that the theory of factor proportions predicted that the capital abundant country exported capital-intensive goods and imported labour-intensive goods, and the labour surplus country did the opposite.

- The **non-availability** explains international trade by the fact that each country imports the goods that are not available at home. This unavailability may be due to lack of natural resources (oil, gold, etc. : this is *absolute* unavailability) or to the fact that the goods cannot be produced domestically, or could only be produced at prohibitive costs (for technological or other reasons): this is *relative* unavailability. On the other hand, each country exports the goods that are available at home.

- The hypothesis was proposed by economist Staffan Burenstam Linder in 1961 as a possible resolution to the Leontief paradox, which questioned the empirical validity of the Heckscher-Ohlin theory (H-O). H-O predicts that patterns of international trade will be determined by the relative factor-endowments of different nations. Those with relatively high levels of capital in relation to labor would be expected to produce capital-intensive goods while those with an abundance of labor relative to (immobile) capital would be expected to produce labor intensive goods. H-O and other theories of factor-endowment based trade had dominated the field of international economics until Leontief performed a study empirically rejecting H-O. In fact, Leontief found that the United States (then the most capital abundant nation) exported primarily labor-intensive goods. Linder proposed an alternative theory of trade that was consistent with Leontief's findings. The Linder hypothesis presents a demand based theory of trade in contrast to the usual supply based theories involving factor endowments. Linder hypothesized that nations with similar demands would develop similar industries. These nations would then trade with each other in similar, but differentiated goods.

4.5 Key-Words

1. Labour surplus : Surplus labour is a concept used by Karl Marx in his critique of political economy. It means labour performed in excess of the labour necessary to produce the means of livelihood of the worker ("necessary labour"). According to Marxian economics, surplus labour is usually "unpaid labour". Marxian economics regards surplus labour as the ultimate source of capitalist profits.

2. Factor price equalization : Factor price equalization is an economic theory, by Paul A. Samuelson (1948), which states that the prices of identical factors of production, such as the wage rate, or the return to capital, will be equalized across countries as a result of international trade in commodities. The theorem assumes that there are two goods and two factors of production, for example capital and labour. Other key assumptions of the theorem are that each country faces the same commodity prices, because of free trade in commodities, uses the same technology for production, and produces both goods. Crucially these assumptions result in factor prices being equalized across countries without the need for factor mobility, such as migration of labor or capital flows.

4.6 Review Questions

1. Discuss the modern theories of international trade.

2. Explain the theorem of factor price equalization .

4.7 Further Readings

1. James E. Anderson (2008). "International Trade Theory," The New Palgrave Dictionary of Economics, 2nd Edition.

2. Devashish Mitra, 2008. "trade policy, political economy of," The New Palgrave Dictionary of Economics, 2nd Edition.

3. A. Venables (2001), "International Trade: Economic Integration," International Encyclopedia of the Social & Behavioral Sciences, pp. 7843-7848.

4. Maurice Obstfeld (2008). "international finance," The New Palgrave Dictionary of Economics, 2nd Edition.

5. Giancarlo Corsetti (2008). "new open economy macroeconomics," The New Palgrave Dictionary of Economics, 2nd Edition.

6. Reuven Glick (2008). "macroeconomic effects of international trade," The New Palgrave Dictionary of Economics, 2nd Edition. Abstract.

7. As at the JEL classification codes, JEL: F51-F55. Links to article-abstract examples for each subclassification are at JEL Classification Codes GuideJEL:F5 links.

Unit 5 : Role of Dynamic Factors : Tastes, Technology and Factors Endowments in Trade

---

CONTENTS

Objectives

Introduction

Objectives

After reading this Unit students will be able to:

• Understand Dynamic Factors.

• Explain Tastes, Technology and Factors Endowments in Trade.

Introduction

So far, we have assumed that each nation has given and unchanging factor endowments and technology (hence a given production possibilities curve) and given and unchanging tastes (hence a given community indifference map). On this premise, we examined the basis and the gains from trade. However, over time a nation's factor endowments change, and its technology may improve. These changes cause its production possibilities curve to shift. Similarly, a nation's tastes may change and result in a different indifference map. All of these changes affect the terms of trade and the volume of trade. How much these are altered depends on the actual type and degree of the changes occurring.

5.1 Growth in Factor Supplies Through Time

If technology remains the same but the factors of production available to a nation increase, the nation's production possibilities curve shifts outward. This shift is uniform or symmetrical (so that the new production possibilities curve has the same shape as the old one) if labor and capital grow in the same proportion. This is called *balanced growth*. If only Ihe nation's supply of labor increases or if its supply of labor increases proportionately more than its supply of capital, then the nation's production possibilities curve shifts more along the axis measuring the L-intensive commodity than along the axis measuring the K-intensive commodity. The opposite shift occurs if only the nation's supply of capital increases or if its supply of capital increase proportionately more than its supply of labor.

According to the *Rybczynski theorem*, at constant relative commodity prices, the growth of only one factor leads to an absolute expansion in the output of the commodity using the growing factor intensively and to an absolute reduction in the output of the commodity using the nongrowing factor intensively.

Example 1 : If only the supply of labor increases in the U.K. or if the supply of labor increases proportionately more man the supply of capital and its technology remains the same, then the U.K.'s production possibilities curve or transformation curve might shift outward from TT to T'T' as shown in Figure 5.1. Note that the shift is greater along the horizontal axis, which measures cloth (the L-intensive commodity), than along the vertical axis, which measures wheat (the K-intensive commodity). The shift along the cloth axis in Figure 5.1 is exaggerated for pedagogical reasons. Even if the supply of labor alone increases in the U.K., the U.K. production possibilities curves will nevertheless shift slightly upward since tabor is also used in the production of wheat (the K-intensive commodity). However, according to the Rybezynski theorem, the output of cloth would rise while the output of wheat would fall in the U.K. at constant P_C/P_w . For the effect of other types of factor changes on the production possibilities curve of the U.K. and the U.S. and for graphical illustrations and an intuitive proof of the Rybezynski theorem.

5.2 Technical Progress

Technical progress increases the productivity of a nation's factors of production and has the same general effect on the nation's productions possibilities curve as an increase in the supply of its factors. There are at least three types of technical progress :
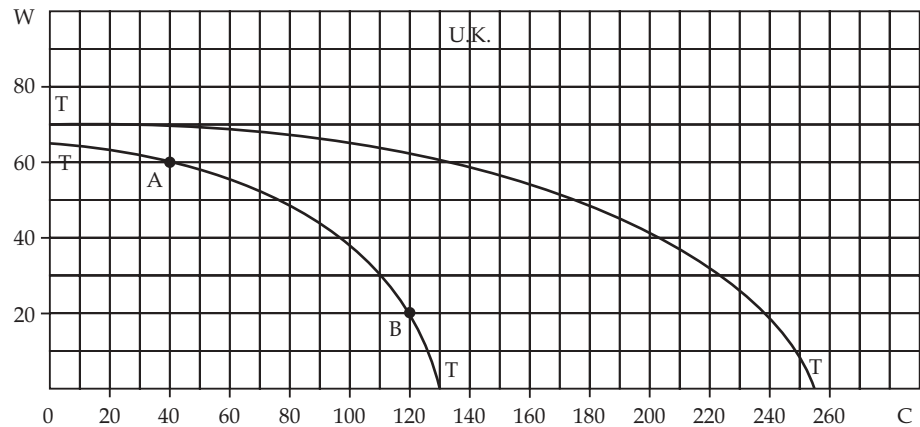
Figure 5.1

(1) *K-Saving technical progress* throughout the economy increases the productivity of labor proportionately more than that of capital. As a result, L is substituted for K in production at constant *w/r*, and K/L falls in the production of both commodities. This means that a given output can now be produced with fewer units of L and K but also with lower K/L (higher L/K). K-saving technical progress is equivalent to a proportionately greater increase in the supply of labor than of capital (with unchanged technology). For example, K-saving technical progress in the U.K. equally applicable to cloth and wheat production might cause an outward shift in the U.K. production possibilities curve torn TT to T'T' in Figure 5.1. Note that technical progress is defined at constant *w/r*, and constant returns to scale are assumed in production.

(2) *L-saving technical progress* is exactly the opposite of K-saving technical progress.

(3) *Neutral technical progress* increases the productivity of L and K by the same proportion and results in a uniform or symmetrical outward shift in the nation's production possibilities curve.

5.3 Change in Factor Supplies, Technology and Trade

When there is an increase in a nation's supply of factors of production and/or technical progress, the nation's production possibilities curve shifts outward. With unchanged tastes, this causes a change

in the terms of trade, in the volume of trade and in the distribution of the gains from trade between the two nations. The actual result depends on the type and degree of the changes occurring.

Example 2 : In panel A of Figure 5.2, we see that before any change in factor endowments and/or technology the U.K. produces at point B on TT, exports 60C for 60W and consumes at point E on III. . This gives point E on the U.K. offer curve in panel B. With unchanged tastes at the same terms of trade of P_B but after TT shifts to T'T', the U.K. would like to produce at point M on T'T', export 150C for 150W and consume at point U on VII. This gives point U on offer curve U.K.* in panel B. If nothing changes in the U.S. (so that the U.S. offer curve in panel B of Figure 5.2, we see that at P_B the U.K. would like to export more of its cloth and import more American wheat than the U.S. is willing to trade at to P_B (see panel B of Figure 5.2). Thus, the terms of trade move against the U.K. from to P_B to P_G. At P_G , the U.K. produces at point N on T'T, exports 140C for 70W and consumes at point R on V (see panel A of Figure 5.2). This corresponds to the new equilibrium point R in panel B, where the U.K. and the U.S. offer curves cross. Thus, the U.S. shares in the benefit of the growth taking place in the U.K.
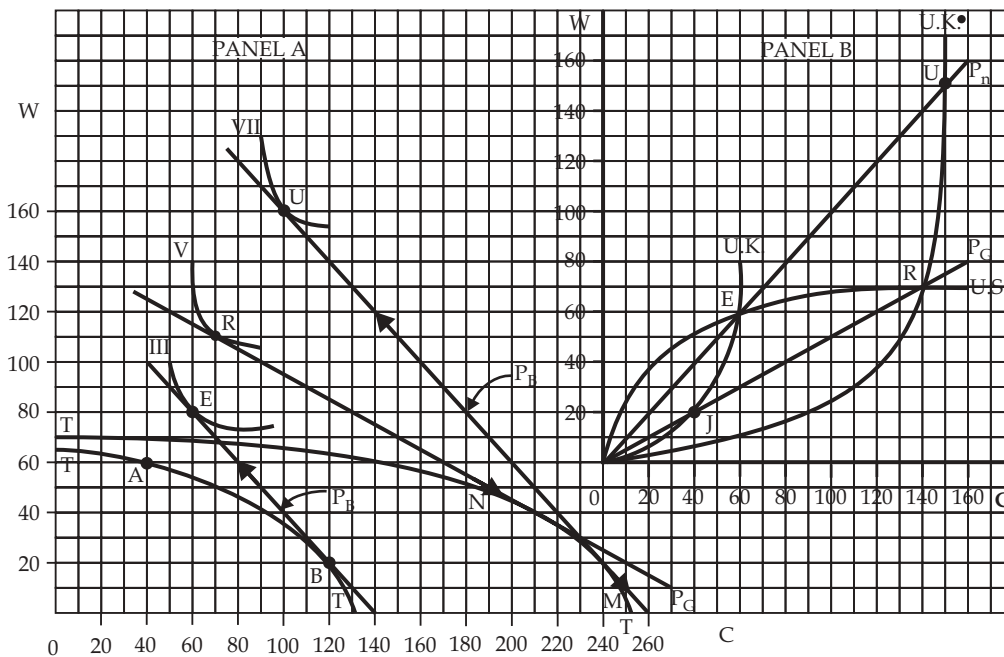
Figure 5.2

The general rule is that if at unchanged terms of trade the nation wants to trade more after growth, its terms of trade partner. If the nation wants to trade less, it not only retains all of the benefits of its growth but is also likely to gain from better terms of trade.

5.4 Change in Tastes and Trade

A nation's *offer curve* also shifts if the nation's supply of factors and technology remain unchanged but its tastes change. Thus, a change in tastes also alters the volume of trade and the nation's terms of trade. More specifically, if a nation's tastes shift away from its *importable commodity* and toward its *exportable commodity* (other things equal), the volume of trade declines and the nation's terms of trade improve. The reserve occurs with an opposite change in the nation's tastes. However, since the nation's indifference map is altered by the change in tastes, we can no longer determine how the nation's welfare is affected.

Example 3 : If the U.K.'s tastes shift away from wheat (its importable commodity) and toward cloth (its exportable commodity—other things being equal), offer curve U.K. in Figure 5.3 shifts up or rotates counterclockwise, say, to offer curve U.K.' This is because the U.K. now wants wheat less

intensely and is willing to offer less of its cloth than before for each quantity of wheat imported. Thus, with offer curve U.K.' the U.K. exchanges only 20C for 40W at the new equilibrium point J' and the terms of trade of the U.K. improve to $P_G' = 2$ (from $P_B = 1$). This improvement in the terms of trade, by itself, tends to improve the welfare of the U.K. The reduction of specialization in production and in the volume of trade, by itself, tends to reduce the welfare of the U.K. (from the previous trade position before the change in tastes).
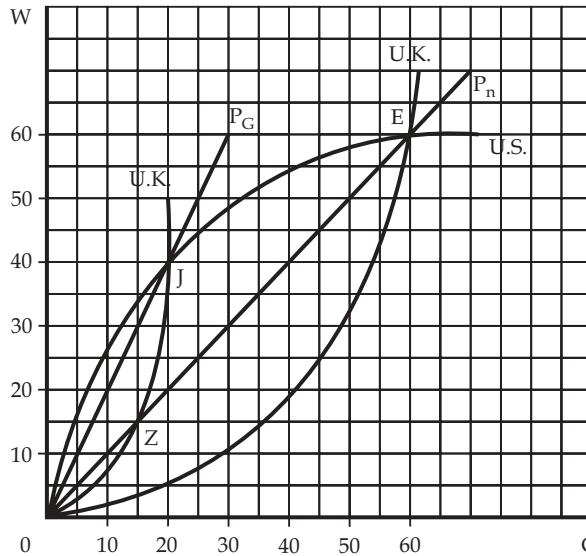
Figure 5.3

Whether on balance the welfare of the U.K. improves or not depends on the relative strength of these two opposing forces. However, we can no longer use commodity indifference curves to answer this question because when tastes change in the U.K., the entire indifference map of the U.K. changes and indifference curves cross.

If, on the other hand, the U.K.'s tastes shift from cloth to wheat, the U.K. offer curve will rotate clockwise and result in a greater volume of trade but reduced terms of trade for the U.K. For the effect of changes in U.S. tastes, separately and at the same time as changes in the U.K.'s tastes occur.

5.5 Dynamic Factors, Trade and Development

With the exception of a handful of nations in North America and is Western Europe and Japan, most nations of the world are classified as less developed countries (LDCs) or, to put it more positively, as developing countries. LDCs presently account for less one-fourth of world trade. Aside from a small group of *newly industrializing economies* (NIEs) (especially South Korea, Singapore, Taiwan and Hong Kong) which are growing very rapidly based on the export of manufactured goods, most of the trade of other LDCs involves the export of raw materials, fuels, minerals and some food products to the industrialized, rich and developed countries (DCs), mostly in exchange for manufactured goods. LDCs complain that because of this trade pattern, because their internal conditions differ widely from those in DCs and because of the way in which the present international monetary system operates, most of the benefit of their own growth accrues to DCs, primarily in the form of secularly improving terms of trade. Thus, trade can no longer serve as an *engine of growth* for today's developing countries as it did for the *regions of recent settlement* (the U.S., Canada, Argentina, Uruguay, Australia, New Zealand, and South Africa) during the nineteenth century.

LDCs advocate a *new international economic order* (NIEO), which involves higher prices for their traditional exports, schemes to stabilize their export proceeds, increased foreign aid, preferential

treatment for their manufactured exports to DCs and reforms of the international monetary system that would take their interests into consideration.

For a discussion of multinational corporations and international labor migration.

5.6 Factors Endowment in Trade

In economics a country's factor endowment is commonly understood as the amount of land, labor, capital, and entrepreneurship that a country possesses and can exploit for manufacturing. Countries with a large endowment of resources tend to be more prosperous than those with a small endowment, all other things being equal. The development of sound institutions to access and equitably distribute these resources, however, is necessary in order for a country to obtain the greatest benefit from its factor endowment. The Heckscher-Ohlin model (H-O model) is a general equilibrium mathematical model of international trade, developed by Eli Heckscher and Bertil Ohlin at the Stockholm School of Economics. The model essentially says that countries will export products that use their abundant and cheap factor(s) of production and import products that use the countries' scarce factor(s).

Nonetheless, the New World economies inherited attractive endowments such as conducive soils, ideal weather conditions, and suitable size and sparse populations that eventually came under the control of institutionalizing European colonists who had a marginal economic interest to exploit and benefit from these new discoveries. Colonists were driven to yield high profits and power by reproducing such economies' vulnerable legal and political framework, which ultimately led them towards the paths of economic developments with various degrees of inequality in human capital, wealth, and political power.

Factor endowments in the New World

A classical example often cited to emphasize the importance of institutions in developing a country's factor endowment is that of North America (the United States and Canada) around the turn of the 19th to 20th century. It is commonly argued that these countries benefited greatly by borrowing many of Britain's institutions and laws. While North America undoubtedly gained from this borrowing, this does not fully explain why the rest of the New World (which also enjoyed a large factor endowment and access to British institutions) did not develop in a similar way. In fact, data shows that connection between the prosperity of the colonizing and the wealth of the colony was weaker than many thought. The future United States and Canada surpassed several British established colonies in the Caribbean, such as Barbados, Jamaica, Belize, and Guyana. In fact, the United States converged on the world economic leader, measured in GDP/capita, the UK. In 1910, the United States overtook the UK and began to diverge from it until about the 1950s. This shows that there must have been another explanation as to why the future United States and Canada developed at a faster rate than other colonies in the region.

Did u know? Cuba and Brazil primarily grew lucrative products such as cotton, coffee, and sugar, which required hand picking and most efficiently picked when picked by hands in unison, whereas the United States was generally a wheat producer.

Kenneth Sokoloff and Stanley Engerman argue in their article "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World" that the difference between North America and the rest of the New World was not just in institutions but in the nature of their respective factor endowments. Countries like Brazil and Cuba had an extremely large yet concentrated factor endowment that tended toward exploitation, a hierarchical social system and exhibited economies of scale. The true advantage of the United States and Canada lay in a more equitable distribution of factors that could not be exploited on an extremely large scale. This distribution led to a more open and opportunistic economy, and eventually to long-term prosperity. For example, because wealth and power were distributed relatively equally in the United States and in Canada, these two countries

led the rest of the Americas in providing education on a broader scale. Education is an important factor to improve technology in order to boost productivity, which is the reason that US and Canada surpassed the others. Greater access to education allowed for greater investment in human capital, which increases productivity and contributed to the United States and Canada's economic growth. According to Sokoloff and Engerman's article "History Lessons: Institutions, Factors Endowments, and Paths of Development in the New World," not only the United States had relatively equal distribution of wealth, it had relatively homogeneous population, political power and human capital. United States and Canada's relatively equal distribution of wealth, amount of human capital and political power utimately affected development of institution, extent of franchise, and public education that persist and influence growth of the country. The open franchise brought elimination of wealth and literary requirement by 1940 in Canada and the United States (literacy requirement was still enforced in US southern regions only). Again, the open franchise was possible because the United States endowed a land suitable for wheat growing thus had a large body of middle class unlike Brazil and Cuba where they exhibited small elites, some overseers and large slave population. United States, then, outgrows other New World countries and eventually diverged from Cuba and Brazil in the late 18th and early 19th centuries.

> *Notes* The open franchise in the United States and Canada was possible due to the large voting body of middle class and small elites.

Sokoloff and Engerman hypothesize that in societies founded with greater inequality, the elites gained more power to influence the choice of legal and economic institutions. In those countries which are inequal, small elites restrict majority people's rights, such as education and votes, to perpetuate the social structures and continue to make themselves "elites." The U.S. began its economic growth largely through slave labor and trade of the output of that labor. As the elites enacted policy to generate more economic equality, for example by increasing literacy rates, the U.S. GDP per capita pulled ahead of other long-since established countries along with the literacy rates. It is essential to note that factor endowments played a crucial role in shaping the colonies institutions and economic growth; colonies with a richer quality of soil grew cash crops such as sugar, coffee, and cotton, which were most efficiently grown using plantation systems. As such, the demand for not only slave labor but also peonage within these colonies grew. Due to the vast inequality that the society developed due to a small elite population in comparison to the vast laborer population, they were able to maintain the wealth and power within the elite class via establishing a guarded franchise. The inequalities within the cash crop colonies resulted in their economy not being able to expand and grow as fast as the U.S and Canada, due to restrictive policies. Those policies in inequal countries curb the intellectual development of most people who are only required to do simple manual jobs; however, US and Canada encourage their people to take part in education. As a result, US and Canada excel with higher productivities which are supported by advanced technology.

Self-Assessment

1. **Choose the correct options:**

   (*i*) Dynamic factors in trade theory refer to changes in

     (*a*) factor endowments,   (*b*) technology,

     (*c*) tastes,   (*d*) all of the above.

   (*ii*) If a nation's technology remains unchanged but its supplies of labor and capital grow in the same proportion, then the nation's production possibilities curve shifts outward

     (*a*) evenly along its entire length,

     (*b*) more along the axis measuring the L-intensive commodity,

     (*c*) more along the axis measuring the K-intensive commodity,

     (*d*) any of the above is possible.

(*iii*) If a nation's technology remains unchanged but its supply of labor grows proportionately more than its supply of capital, then the nation's production possibilities curve shifts outward

    (*a*) evenly along its entire length,

    (*b*) more along the axis measuring the L-intensive commodity,

    (*c*) more along the axis measuring the K-intensive commodity,

    (*d*) any of the above is possible.

(*iv*) If a nation's supply of labor increases but its supply of capital and technology remains unchanged, then the nation's production possibilities curve shifts outward

    (*a*) only along the axis measuring the L-intensive commodity,

    (*b*) mostly along the axis measuring the L-intensive commodity but also a little along the axis measuring the K-intensive commodity,

    (*c*) mostly along the axis measuring the K-intensive commodity,

    (*d*) any of the above is possible.

(*v*) According to the Rybczynski theorem, the growth of only one factor, at constant relative commodity prices, leads to an absolute expansion in the output of

    (*a*) both commodities,

    (*b*) the commodity using the growing factor intensively,

    (*c*) the commodity using the non growing factor intensively,

    (*d*) any of the above.

5.7 Summary

- If technology remains the same but the factors of production available to a nation increase, the nation's production possibilities curve shifts outward. This shift is uniform or symmetrical (so that the new production possibilities curve has the same shape as the old one) if labor and capital grow in the same proportion.

- Technical progress increases the productivity of a nation's factors of production and has the same general effect on the nation's productions possibilities curve as an increase in the supply of its factors.

- With unchanged tastes, this causes a change in the terms of trade, in the volume of trade and in the distribution of the gains from trade between the two nations. The actual result depends on the type and degree of the changes occurring.

- A nation's *offer curve* also shifts if the nation's supply of factors and technology remain unchanged but its tastes change. Thus, a change in tastes also alters the volume of trade and the nation's terms of trade.

- With the exception of a handful of nations in North America and is Western Europe and Japan, most nations of the world are classified as less developed countries (LDCs) or, to put it more positively, as developing countries. LDCs presently account for less one-fourth of world trade.

5.8 Key-Words

1. Balanced growth	:	Equal rates of factor growth and technical progress in the production of both commodities.
2. Engine of growth	:	The driving force behind economic growth. In the nineteenth century, for example, exports were the leading sector that propelled the economies of the regions of recent settlement into rapid growth and development.
3. Exportable commodity	:	A commodity (such as aircrafts in the U.S.) that a nation produced for domestic use and for export.

4. Importable commodity : A commodity (such as cars in the U.S.) that a nation produces at home and also imports (because of incomplete specialization in production).

5. K-saving technical progress : Technical progress that increases the productivity of labor proportionately more than the productivity of capital and results in an increase in L/K at constant relative factor prices.

6. L-saving technical progress : Technical progress that increases the productivity of capital proportionately more than the productivity of labor and results in an increase in K/L at constant relative factor prices.

7. Neutral technical progress : Technical progress that increases the productivity of labor and capital in the same proportion so that K/L remains the same at constant relative factor prices.

5.9 Review Questions

1. What is the role of dynamic factors? Discuss.
2. What are the changes in factor supply and technology and trade? Give examples.
3. Write a short note on the dynamic factors, trade and development.

Answers: Self-Assessment

1. (*i*) (*d*)   (*ii*) (*a*)   (*iii*) (*b*)   (*iv*) (*b*)   (*v*) (*b*)

5.10 Further Readings

Books

1. Kenneth L. Sokoloff, Stanley L. Engerman. "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World". The Journal of Economic Perspectives Vol 14 No.3 (2000): pp. 217-232

2. Kenneth L. Sokoloff, Stanley L. Engerman. "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World". The Journal of Economic Perspectives Vol 14 No.3 (2000): pp. 217-232

3. Kenneth L. Sokoloff, Stanley L. Engerman. "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World". The Journal of Economic Perspectives Vol 14 No.3 (2000): pp. 217-232

4. Kenneth L. Sokoloff, Stanley L. Engerman. "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World". The Journal of Economic Perspectives Vol 14 No.3 (2000): pp. 217-232

Unit 6 : Rybczynski Theorem

Objectives

After reading this Unit students will be able to:

•   Define Rybczynski Theorem.

•   Explain Rybczynski Theorem.

Introduction

The Rybczynski theorem was developed in 1955 by the Polish-born English economist Tadeusz Rybczynski (1923-1998). The theorem states: At constant relative goods prices, a rise in the endowment of one factor will lead to a more than proportional expansion of the output in the sector which uses that factor intensively, and an absolute decline of the output of the other good.

In the context of the Heckscher-Ohlin model of international trade, open trade between regions means changes in relative factor supplies between regions, that can lead to an adjustment in quantities and types of outputs between those regions, that would return the system toward equality of production input prices like wages across countries (the state of factor price equalization).

It states that at constant prices, an increase in one factor endowment will increase by a greater proportion the output of the good intensive in that factor and will reduce the output of the other good. An increase in the supply of labour expands production possibilities disproportionately in the direction of the production of labour-intensive good (wheat), while an increase in the supply of capital expands them disproportionately in the direction of the production of capital- intensive good (cloth).

Suppose the supply of capital increases by 10% and that of labour is unchanged. If both goods continue to be produced, then factor prices will not change (because of factor-price equalisation theorem) and so the techniques of production will also not change.

As a result of increase in capital,

1. the output of both goods cannot rise by 10% because this would require 10% more labour, and the supply of labour has not changed;

2. output of both goods cannot rise by more than 10%,

3. output of both goods cannot fail to rise by 10% because otherwise the increased capital could not all be utilised;

4. thus the output of one rises by more than 10% and that of the other does not. Because cloth is capital intensive, it must be cloth output that rises more than 10%. The labour supply has not changed, but the cloth industry has expanded and so has increased Us use of labour. Therefore, the output of wheat must actually fall.

By combining this result with the Heckscher-Ohlin theorem, we can see how economic growth affects a nation's trade. If a country's capital increases by 10%, national income will rise by some smaller proportion, because only part of national income comes from the earnings of capital. This increased income will normally be spent on both goods, so that at constant prices, national demand for both goods will rise by less than 10%.

Did u know?   Rybczynski Theorem discusses the effect of economic growth on a nation's trade.

According to Rybczynski Theorem, the supply of capital-intensive good (cloth) rises more than 10%, while the supply of labour-intensive good (wheat) falls.

Thus, cloth supply rises relative to demand, and wheat demand rises relative to supply. Now, if the country is capital intensive, then according to the Heckscher-Ohlin theory, it exports cloth and imports wheat, so that the growth of capital causes the country to trade more at each price.

Thus, its offer curve shifts outward. If the country is labour abundant, its offer curve shifts inward. The general conclusion is economic growth that accentuates country's relative factor abundance shifts its offer curve it; economic growth that moderates the country's relative factor abundance shifts its offer curve in.

6.1 Rybczynski Theorem

The Rybczynski theorem displays how changes in an endowment affects the outputs of the goods when full employment is sustained. The theorem is useful in analyzing the effects of capital investment, immigration and emigration within the context of a Heckscher-Ohlin model. Consider the diagram below, depicting a labour constraint in red and a capital constraint in blue. Suppose production occurs initially on the production possibility frontier (PPF) at point A.
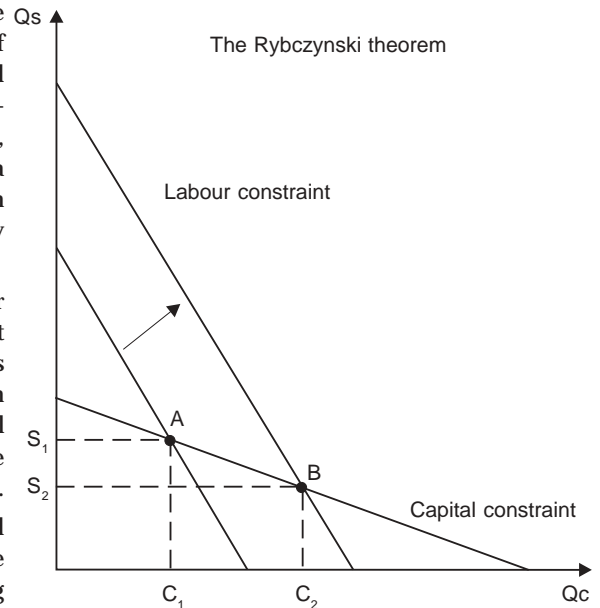
Suppose there is an increase in the labour endowment. This will cause an outward shift in the labour constraint. The PPF and thus production will shift to point B. Production of clothing, the labour intensive good, will rise from C1 to C2. Production of cars, the capital-intensive good, will fall from S1 to S2.

If the endowment of capital rose the capital constraint would shift out causing an increase in car production and a decrease in clothing production. Since the labour constraint is steeper than the capital constraint, cars are capital-intensive and clothing is labor-intensive.

Figure 6.1

In general, an increase in a country's endowment of a factor will cause an increase in output of the good which uses that factor intensively, and a decrease in the output of the other good.

The factor price equalization theorem assumes no change in factor supplies. We will now see what happens to the trading relationships if in one of the two countries there is an increase in the labour force or an increase in the capital stock. This is the essential basis of the so called Rybczynski theorem.

Let us start by assuming that in one of the two countries there is increase in labour supply. The country's original factor endowments are measured by the box ACBD in Diagram 1. The point of origin for cloth (the labour intensive good) is A, and the point of origin for steel (the capital intensive

good) is at B and F. Cloth is the exported good while steel is the import-competing good. Once, there is increase in the supply of labour, we will have a new box of the size of AEFD, where CE measures the increase in labour force.

Initially, the country produces at point P on its contract curve APB. The factor intensity in the production of cloth, will then be AP, and the factor intensity in the production of steel will be BP. When the country gets an increase in labour supply (equal to CE) with no change in capital stock, the country would move towards a new production point, R on its new contract curve ARF. What are the effects of this movement on factor prices, product prices, production levels, terms of trade, etc ?
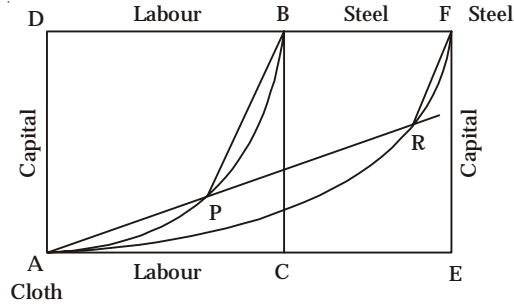
Figure 6.2: Increase in Labour Supply : Rybczynski case.

You will notice, first of all, that there is no change in factor prices. The factor intensities in the production of the two goods remain unchanged. Originally, the factor intensity in cloth production was PA and now it is RA; in the production of steel, it was BP earlier, and now it is FR (FR is parallel to BP). This means that factor price ratio has remained unchanged. (This would prevent factor price equalization from taking place if the labour abundant country experiences labour force increase and the capital abundant country experiences increase in its capital stock). Since factor prices are kept constant, the commodity prices also would remain constant. In moving from point P to R, there is no change in either factor prices or product prices, and the only change is in the volumes of the two goods produced.

The change in the volume of output produced in moving from point P to R can be measured by the distance from point A to point P and to point R (in the case of cloth) and from point B to P and from point F to R (in the case of steel). The distance FR is shorter than the distance BP, which means that less of steel is being produced at R than at P. In the same way, the distance AR is more than the distance AP, which must, therefore, mean that more of cloth is being produced at R as compared to P. All this would, then, mean that if the supply of one of the two factors is increased while the other is kept constant, the production of the good intensive in the increasing factor will increase in absolute amounts, whereas the production of the other good would decrease absolutely. The commodity and factor prices remain constant.

Notes The Rybczynski theorem has several implications to welfare, terms of trade, commodity and factor prices, etc. But its implications to factor-price equalization theorem is rather clear cut, i.e, if the supplies of the abundant factor of production expands rapidly, it could keep the factor price ratio in the country constant, preventing factor-price equalization among countries from taking place.

In this case, what has happened, is as follows. The amount of labour has increased, and all the new labour has gone into the labour-intensive industry (viz. cloth). From the capital intensive industry (viz. steel) not only capital but also some labour has moved out into the labour intensive industry. This is reflected in the fact that the production of cloth has expanded, while that of steel has contracted. If the labour force continues to expand indefinitely, the country would soon become completely specialized in the production of cloth.

To continue the argument it is necessary to use another geometric illustration. In Figure 6.2, the production-possibility curve TT is derived from the box ACBD in Figure 6.3. The international terms of trade are given by the line P_0P_0, which is tangent to TT at point S. What about the new production-possibility curve ? What shape will it have ? All, we can say, is, that a price line such as P_1P_1, Which is parallel to P_0P_0, must be tangent to the new production-possibility curve below the line SN. This is, because, as we have just seen, at constant prices more cloth (exportables) will be produced and less steel (importables) will be produced. So, if we derive the new production-possibility curve FF from the box AEFD in Figure 6.2, it will appear as it does in Figure 6.2.
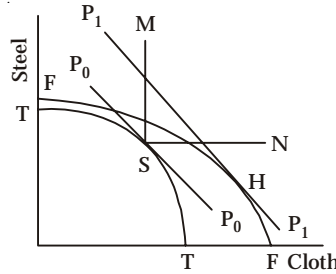
Figure 6.3: Effects of Factor Growth on Trade Equilibrium.

Point H cannot be an equilibrium point unless steel is an inferior good. The increase in labour force and a consequent shift in the production-possibility curve would imply an increase in the national income. Barring inferior goods, the demand for both goods must increase. Therefore, the new equilibrium point will have to be somewhere on the new production-possibility curve FF between where it is cut by the lines SN and SM. The slope of this segment on the FF curve is not as steep as the slope of the TT line at point S. That implies that the relative price of steel will be higher in the new equilibrium situation, which, in other words, means that an increase in labour supplied will lead to a deterioration in the country's terms of trade.

The factor growth, therefore, would lead to a deterioration in the terms of trade for that good which uses the increasing factor intensively. In this case, labour supply increased, and the country's export good was the labour-intensive good, cloth. The cloth price would decrease turning the terms of trade against the country. All this is based on the premise that neither of the two goods is an inferior good.

Self-Assessment

1. **Choose the correct option:**

    (*i*) The Rybczynski theorem was developed in .... by the Polish-born English economist Tadeusz Rybczynski

    (*a*) 1955   (*b*) 1950

    (*c*) 1970   (*d*) 1960

    (*ii*) This econometric find was the result of Professor .... attempt to test the Heckscher-Ohlin theory empirically.

    (*a*) Tadeusz Rybczynski   (*b*) Wassily W. Leontief's

    (*c*) Heckscher-Ohlin   (*d*) None of these

    (*iii*) In ...., Leontief found that the U.S. (the most capital-abundant country in the world) exported labor-intensive commodities and imported capital-intensive commodities, in contradiction with Heckscher-Ohlin theory ("H-O theory").

    (*a*) 1954   (*b*) 1950

    (*c*) 1980   (*d*) None of these

    (*iv*) A 1999 survey of the econometric literature by .... concluded that the paradox persists, but some studies in non-US trade were instead consistent with the H-O theory.

    (*a*) Tadeusz Rybczynski   (*b*) Richard Freeman

    (*c*) Elhanan Helpman   (*d*) None of these

6.2 Summary

- According to the Rybczynski (1955) theorem, applied to the standard two-country, two-good, two-factor model, an increase in one factor will result in an absolute rise in the output of the commodity which is relatively intensive in the increased facgor, and to an absolute fall in the output of the other commodity.

- The generalization of the theorem by Jones (1965) states that "if factor endowments expand at different rates, the commodity intensive in the use of the fastest growing factor expands at a greater rate than either factor, and the other commodity grows (if at all) at a slower rate than either factor."

- The application of Jones' version of the theorem to a model with three goods, one of which is non-traded, two factors and indecomposable inter-industry flows is studied here.

- The introduction of inter-industry flows makes necessary a distinction between net and gross Rybezynski output effects and also between direct and total factor intensities of commodities. It is found that a sufficient condition for the generalized Rybcynski theorem, defined in terms of total factor intensities, to hold for both net and gross outputs, net output changes being proportionately greater than gross changes, is that the net output change of the non-traded good is bounded by the factor changes.

- This result is compared with earlier findings and the meaning of the sufficient condition is discussed in terms of basic demand parameters.

6.3 Key-Words

1. New international economic order (NIEO) : The demands made developing nations as a group at the United Nations for the removal of alleged injustices in the operation of the present international economic system and for the implementation of specific measures to facilitate the development of these nations.

2. Newly industrializing economies (NIEs) : Economies such as South Korea, Singapore, Taiwan and Hong Kong which are growing very rapidly based mostly on export growth.

3. Regions of recent settlement : The mostly empty and resource-rich lands, such as the U.S., Canada, Argentina, Uruguay, Australia, New Zealand and South Africa, that Europeans settle during the nineteenth century.

4. Rybczynski theorem : A theorem postulating that at constant relative commodity prices the growth of only one factor leads to an absolute expansion in the output of the commodity using the growing factor intensively and to an absolute reduction in the output of the commodity using the non growing factor intensively.

6.4 Review Questions

1. What do you mean by the Rybczynski Theorem? Discuss.

2. Illustrate the diagram with the effects of factor growth on trade equilibrium.

Answers: Self-Assessment

1. (*i*) (*a*)   (*ii*) (*b*)   (*iii*) (*a*)   (*iv*) (*c*)

6.5 Further Readings

Books

1. Altonji, Joseph, and David Card. 1991. "The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives." In John Abowd and Richard Freeman, eds, Immigration, Trade, and the Labor Market, (Chicago, IL: University of Chicago Press), 201-234.

2. Bernard, Andrew B. and J. Bradford Jensen. 1999. "Understanding Increasing and Decreasing Wage Inequality." In Robert C. Feenstra, ed., The Impact of International Trade on Wages, Chicago: University of Chicago Press.

3. Bernstein, Jeffrey R. and David E. Weinstein. 1998. "Do Endowments Predict the Location of Production? Evidence from National and International Data." NBER Working Paper No. 6815.

4. Blackorby, Charles, William Schworm, and Anthony Venables. 1993. "Necessary and Sufficient Conditions for Factor Price-Equalization." Review of Economic Studies 60: 413-434.

Dilfraz Singh, Lovely Professional University

Unit 7 : Causes of Emergence and Measurement of Intra-Industry Trade and Its Impact on Developing Economics

---

CONTENTS

Objectives

Introduction

Objectives

After reading this Unit students will be able to:

* Discuss the Causes of Emergence and Measurement of Intra-Industry Trade.

* Explain the Balassa and Grubel-Lloyd Indices.

Introduction

Having earlier noted that in the European Common Market "much of the increased trade in manufactures occurred within rather than between commodity groups" . An explanation of this phenomenon in terms of product differentiation in consumer goods, machinery, transport equipment, and intermediate products and introduced statistical indicators to measure the extent of intra-industry specialization.

Subsequently, additional evidence was provided on the extent of intra-industry specialization in manufactured goods among the EEC countries. For purposes of empirical measurement, an industry has been defined to include commodities that have high substitution elasticities in production. In practice, limitations of data availability have led to the use of a 91 industry classification scheme consisting of 3-digit and 4-digit items in the U.N. Standard International Trade Classification (SITC), and combinations thereof.

The use of a technological definition of an industry is not open to the strictures Lipsey (1976) addressed to subsequent work by Grubel and Lloyd (1975), who employed 3-digit SITC categories in the calculations regardless of the technological characteristics of the product within each category.

And although a further disaggregation of the data would be desirable in particular instances, Hesse (1974) and Willmore (1974), respectively, have shown that a high degree of intra-industry specialization is apparent in the European Common Market and the Central American Common Market, even if a very dis-aggregated commodity classification scheme is employed. At any rate, recognizing the limitations of the use of the statistical indicators to measure the extent of intra-industry specialization

in a single country at a particular point of time, the author has used these indicators to make comparisons over time and among countries.

The purpose of this section is to discuss the methodological aspects of the measurement of intra-industry trade, rather than a documentary study of intra-industry trade applied to a country, or group of countries. Most of the references in this section, as well as the section below on the determinants of intra-industry trade, contain documentation of intra-industry trade levels for the countries under study that are too numerous to mention here in this review.

But before we can discuss any measurement of intra-industry trade, we must decide what we are to measure. This is not a philosophical question, but a practical one as the history of empirical intra-industry trade has been mired by allegations of being a "statistical phenomenon" (Lipsey, 1976). The charge of being a statistical phenomenon is not an idle one. At the 3-digit SITC level of aggregation, canoes and 200,000 tonne tankers are in the same "ships and boats" industry; at the same level of aggregation, table model radios and airport flight control equipment are in the "telecommunications apparatus" industry (Lipsey, 1976). Also, Finger (1975) notes that trade overlap is not inconsistent with classical trade theory if empirical product groups do not correspond with the appropriate factor proportions groupings. Industry/product categories have become sufficiently disaggregated to disregard these early claims of intra-industry trade being a statistical phenomenon. The Harmonized Tariff Schedule (HTS) has a 10-digit classification system with over 20 000 entries that not only separate canoes from 200 000 tonne tankers, but also from any other boat not designed to be used with motors or sails-a classification that is even distinct from a rowboat. However, aggregation should not be forgotten since it may still be an issue for other reasons. Nevertheless, we will now move on to the various measures.

Did u know?   Intra-industry trade is defined as the simultaneous export and import of goods in the same industry.

7.1 Models of Intra-Industry Trade: Basic Theoretical Underpinnings

So far as the historical source is concerned, IIT was identified while trying to find what culminated into the formation of European Economic Community (EEC). Early research by Dreze (1961) and Balassa (1965) found evidence of increasing intra-industry specialization in the decade following the customs union formation. This led to the work of Grubel and Lloyd (1975) where detailed documentary evidence of IIT at the 2 and 3 digit levels of the SITC for all the major industrialized countries have been provided. Although by the mid-1970s some theorizing has been done [for example, Gray (1973), Grubel (1970), Pelzman (1978)], there was no formal theoretical model. In addition, by the mid-1970s there were few serious attempts at econometric explanation [Pagoulatos and Sorensen (1975), McAleese (1979)-to mention a few]. In particular, theoretical work has become fashionable following some of the theoretical formulations by Dixit and Stiglitz (1977) and Lancaster (1979).

Demand and its Importance in the Determination of Trade Structure: Linder Hypothesis

As mentioned, according to the Linder hypothesis, while the composition of trade in primary products may be explained by the factor proportions theory, the pattern of trade in manufactured goods is mainly determined by the demand characteristics of a country. Specifically, Linder argues that the structure of relative prices of industrial goods in each country is determined by the "representative demand", and that income per capita is the most important single factor influencing the representative demand. Linder claims that "the production functions of goods demanded at home are relatively more advantageous" and he justifies this assertion on the reasons based on the unfamiliarity of the producers with foreign markets as compared with the domestic market and on the existence of scale effects.

Many empirical studies have been done to analyse the empirical links between the structure of industrial exports and the level of income per capita. Hufbauer tested the relationship between income per capita and the composition of trade. The story of Linder is a controversial alternative to factor proportions theory. As regards the bilateral trade intensity, Linnemann has made explicit reference to distance variable which is absent in Linder. Johnson suggested that the positive relationship between trade intensity and "Linder variable" (international similarity in per capita GNP) could be the result of the reality of geographical proximity among countries with similar wealth levels. Thus, Linder variable is a surrogate for distance and distance between partners is deterrent to trade. Following Bergstrand (1985) & Linneman (1986), Hanink has developed a gravity model [Bergstrand (1985)] and analyzed Linder's theme as a "spatial interaction model" i.e., model based on mutual attraction between places.

So far as the empirical testing of Linder corollary relating to the commodity composition of trade in manufactured goods is concerned, the statements is "Potential exports and imports are--when they are manufactured--the same products. An actual import product today is a potential export product today and may be an actual export product tomorrow." This means that there would be a similarity between a country's export vector of manufactures and its import vector of manufactures--irrespective of its level of development. This export-import similarity is measured by either, Finger and Kreinin (1979) Export-Import Similarity (EIS) index or, by Allen's Cosine measure (COS).

Notes   In Linder's version, exports of manufactures are an outgrowth of a home production satisfying the home consumption demand

---

A study by Linnemann and Beers (1988) on the commodity composition of exports of a country and of imports of another country shows that Linder thesis of a potentially relatively stronger trade in manufactures between countries of similar level of per capita income is rejected. On the other hand, the potential intensity of trade would generally seem to increase with increasing per capita income of the trading partners. Gray notes the applicability of Linder's proposition to the explanation of IIT and relates it to the theory of differentiated markets in international trade. Gray calls these goods as 'Linder Goods' and these goods are the primary component of the large volume of trade between countries.

Overlapping demands also arise in the context of product variety i.e., the number of goods in a country's basket of imports/exports. Overlapping demands among rich countries can cover both income elastic and income inelastic goods.

Increasing Returns and Scale Economies

Increasing returns to scale provide an additional factor motivating trade where both countries benefit from trade even when they are identical with respect to tastes and technology. This is supply side explanation of models. Such trade cannot be carried on in conditions of perfect competition and equilibrium will require that the firms involved have some degree of market power. The role of scale economies is of particular interest because of their importance to theoretical models. Most genuine IIT consists of two way trade in differentiated products, since with the exception of strategic trade in oligopoly market situations, homogenous goods IIT is believed to represent border or seasonal trade.

The first departure from the standard competitive model is the Marshallian approach in which increasing returns are assumed to be external to the firm and internal to the industry, allowing perfect competition to remain. According to Krugman, the literature did not seem to offer the interaction of increasing returns and comparative advantage as explanations of trade. Ethier cast his approach to the problem in terms of the two-way trade in intermediate goods, providing a formal basis for relating IIT to external economies linked to the world market size. Subsequently in Ethier, he produced a model in which external and internal economies of scale interact to generate IIT starting from the allocation of resources to production and trade. According to a simplistic version of the scale economy thesis, the large nation because of an assured home market will specialize in goods produced with

increasing returns to industry size. A small nation might occasionally develop a scale economy industry; rely on export sales to justify production. So difference in the sizes of economy could have provided the inducement for trade.

Increasing return internal to the firm in industry gives a different situation incompatible with competitive model. For increasing returns external to the firm, costs fall with the size of the industry not with the size of firms comprising it and hence marginal cost pricing would not lead to losses. External economies are not incompatible with the perfectly competitive structure as it will merely lead to distortions leading to net welfare loss from trade and can be corrected by appropriate government intervention (i.e. taxes and subsidies). It is only when the firms are alone enjoying economies of scale (internal to the firms) that the firms realize the advantage of being alone in the market. It is this hostility to the new entrants and/or a tendency towards merger that threatens the validity of traditional trade theories. Empirical evidence, however, shows that most firms experience internal economies of scale as production expands. This prepares the ground for discussing the third approach to modeling.

Imperfect Competition

The 1970s were marked by substantial progress in the theoretical modeling of imperfect competition. Several trade theorists developed models of trade incorporating non-perfectly competitive market structure. The literature divides itself into two distinct categories in their approach: one strand models the role of scale economies as a cause of trade and keeps the issue of market structure out of the way by assuming Chamberlinian monopolistic competition in market structure. The second strand takes imperfect competition as the base and investigates IRS as a cause of imperfect competition. This falls under the purview of 'oligopoly and trade'.

Did u know? Krugman has been awarded Nobel Prize in Economics in 2008 for his major contribution in this field.

Monopolistic Competition Models

Intuitively it would seem that scale economies would increase the payoff to intra-industry specialization and two way trade in any type of commodity and therefore would be positively associated with the degree of importance of scale effects in an industry. However, as discussed by Greenaway and Milner, even though these models all rely on some type of scale effects to generate IIT, it is not necessarily the case that intensity of such effects determine its share of an industry's trade. For example, Helpman and Krugman (1985), in an alternative to the Chamberlinian framework, modeled IIT by monopolistically, competitive producers of single varieties who are constrained to average cost pricing by freedom of entry. In the 1970s, however, two approaches to this problem were developed. The first, identified with the work of Dixit and Stiglitz (1977) and Spence (1976), made the assumption that each consumer has a taste for many different varieties of a product. As there are alternative approaches, Lancaster (1979), for example, assumed a primary demand for 'attributes' of varieties, with consumers differing in their preferred mix of attributes. Product differentiation here takes the form of offering a variety having attributes that differ from those of existing varieties. Since all these models assume different types of differentiation, a brief taxonomy of terms for product differentiation is given below:

(i) Horizontal differentiation: It refers to differentiation by attributes or characteristic and every consumer has his most preferred "package" of characteristics. Within a given "group" (e.g., in automobiles category as compared to apparels) all products will share certain core characteristics the combination of which determines the products' specifications. It is often called locational differentiation (Hotelling 1919, Lancaster (1980), Helpman (1981)). Pseudo differentiation occurs when the core characteristics of all products in the group are identical, but differentiated by brand image.

(ii)   Vertical differentiation: It is broadly consistent with quality differences i.e. availability of alternative quality grades unlike the earlier case of availability of alternative specifications of a product in a given quality grade.

Oligopolistic Markets and IIT

Because a variety of assumptions can be made regarding conjectural variations, equilibrium outcomes can be generated in a wide range of contexts. Scale economies have also been associated with IIT models in their role as a source of barriers to entry. Strategic interaction among oligopolists whose market share is related to internal economies may result in two way trade as in Brander (1981). Further, with regard to 'Rome market effects" of Helpman and Krugman which leads the monopolies to specialise geographically, Rauch (1989) suggests that a monopolist could internalize Ethier's international external effects by organising as a multinational and engaging in intrafirm IIT. This possibility means that unless the externalities are tied to non-tradable inputs, IIT would not necessary be reduced in the presence of increasing returns. Furthermore, Rauch points out those internal economies linked to a specific location by non-traded inputs or transport costs effects must be strong enough to offset the diseconomies of metropolitan agglomeration before they have the effect of reducing IIT. Because these diseconomies rise steeply with city size he argues that it is difficult to envision large countries as having a comparative advantage capable of affecting the international location of manufacturing simply because they are large. Everything depends on assumption regarding basic structural features of the market concerned and the conjectural variation.

Some important insights into international trade have been gained by adopting the Cournot (1838) assumption that imperfectly competitive firms take each others output as given. Papers by Brander (1981), Brander and Krugman (1983), Neven and Phlips (1984) and Krugman (1984) have made this assumption. The Cournot approach has led to the possibility of trade arising purely because imperfectly competitive firms have an incentive to try to gain incremental sales by "reciprocal dumping" in each others home markets.

The model of Brander (1981) envisages an industry consisting of two firms, each in a different country and each acting as a monopolist in autarky. They take the other firm's deliveries to each market as given. There would be an incentive for each firm to sell a little bit in the others home market as long as the price there exceeds marginal cost and it will continue until, with symmetric firms, each firm has a 50% share of each market. This theory of "reciprocal dumping" is related in important ways to the traditional Industrial Organization literature on basing point pricing and cross hauling. What the new models make clear, however, is that despite the waste involved in transporting the same good in two directions, trade can still be beneficial. The major importance of the Cournot approach, however, lies in its versatility and flexibility of discussion of trade policy. As opposed to this model, different models have been developed where products are vertically differentiated and entry considerations are important. Unlike the Falvey Kurzkowski model, product quality is exogenously given. These models are based fundamentally on the work of Gabszewicz, Shaked, Sulton and Thisse (GSST) in a natural oligopoly framework.

The mathematical complexity of the model is beyond the scope of this review. The above theoretical models have empirical basis and much of the intra-industry trade literature is based upon empirical studies relating to the degree of intra industry trade and the associated country and industry features explaining it. The empirical analyses of IIT requires some discussion regarding problems of measurement and the methodological issues to which we turn in the next section. Assessing the importance of these theories in explaining the pattern of trade is essentially an empirical matter.

7.2 Intra-Industry Trade: Recent Empirical Evidences Using A Global Database

We measure the extent of IIT in manufactures for a set of countries and study the country characteristics and interrelationships between the degree of IIT and some country features. The analysis to be carried out will help explaining the source of generation of IIT in developing countries and the relevance of analysis of IIT, both theoretical and empirically, in the context of developing countries. We use Global

Notes

Trade Analysis Project's (GTAP) Version 6 Database (Dimaranan 2006) to compute extent of such trade. GTAP is a Computable General Equilibrium (CGE) trade model with large database suitable for policy analysis (Hertel ed. 1997). (26) This model divides the world economy into several countries and composite regions. The model and database are widely used for analyzing the effects of issues such as trade liberalization and technological changes. The original Version 6 database consists of 57 commodities and 87 regions expressed in U.S. billion dollars. Typically, the database comprises bilateral trade flows between all the regions. Each set of transactions is recorded at both market prices and agent's prices. GTAP model belongs to the class of computable general equilibrium models (CGE) based on the Australian ORANI model (Dixon et. al. 1982).

We retain the region and sector's identifier number so as to keep it convenient to refer to the GSC classification by mentioning the numbers corresponding to the large database. Typically, there are two concordances of GSC2, one with the Commodity Product Classification (CPC) and the other one with the ISIC Revision 3 (UN). (27) In our empirical analysis, from Table 2 it is seen that the share of IIT in total trade is not a negligible percentage for the developing countries at more advanced level of development. Compared to the developed economies, the share is, no doubt, small. But the interesting picture that comes out from our analysis is that the share is substantial as they diversify their production structure to hi-tech goods especially, with the advent of information and communications technology. Manufactures exports were the developing countries most dynamic part of export sectors in the 1970s and 1980s and also in recent decade. With the rapid growth and economic development of the East Asian newly industrializing countries (NICs), Latin American NICs and the South and South East Asian Countries, there has been a significant increase in intra-industry trade (IIT) in the developing economies. A substantial proportion of these countries IIT has been with their major trading partners e.g., the United States, Japan, the EEC, the U.K., i.e. the developed world. The figures for intra-trade suggest that any presumption that LDCs are more likely to have a comparative disadvantage in advanced manufactures relative to industrial countries and advantage relative to developing countries less developed than them is too simple. Some commodities are too widely produced (e.g., clothing, steel, machinery and transport equipment, etc.) to offer scope for such intra-trade. Countries' whole trading patterns are developing although there has been little change in the composition of manufactured goods' imports. The Asian countries are no longer net importers of manufactures, and in Latin America the ratio of exports to imports is approaching a half. It is clear that the diversification into manufactures, and then into different sectors, has gone well beyond early stages of industrialization or exporting for the major exporters. Table 3 shows commodity-wise patterns of comparative advantage as revealed through their direction and composition of global trade. The 'stylized' picture that comes out from the empirical analysis is that the considerable two way trade of developing economies with the developed economies and also with the world can be explained by the level of development, market size, share of manufacturing value added in GDP and/or share of manufacturing exports in total exports and some trade orientation variable measuring, as a proxy, trade policy intervention. These are all country features. The specific products which have the highest levels of IIT are organic chemicals, glass, leather, iron and steel forms, textile yarn, fabrics, in addition to various types of machinery and equipment including vehicles. Goods with high IIT are more 'sophisticated' and these are, mostly, capital intensive and/or investment goods. Changes in the specialization of certain manufactures towards intra-industry production and exchange is a reflection of the growing similarities between the developing economies and the developed counterpart in terms of relative factor endowments, consumers' preference structure, level of development. It may be reasonably expected that the LDCs will continue to evolve up the ladder of comparative advantage and specialize through international division of labour.

As the developing economies diversify their export through increased IIT, the DCs will have opportunities to export to these countries the products of the industries e.g., textiles, leather, etc. This, however, depends on the LDCs ability to identify and adopt new technologies for achieving such competitiveness. Here, the "vertical specialization" and fragmentation of production processes become important. This means that quality differentiation rather than attribute differentiation is the

appropriate product dimension. Consequently, IIT indices may be expected to be lower and more stable where the goods are vertically differentiated rather than horizontally. Furthermore, it can be inferred from our findings that as industrialization led growth and development in the developing economies proceeds, pushing these countries along the development path towards the matured industrial country stage, intra-industry specialization in production and trade in certain manufacturing commodities will play an increasingly important role in manufacturing production and trade.

7.3 Measurement of Intra-Industry Trade

Empirical work on the measurement of intra-industry trade began in the mid-1960s with Balassa (1966) and the most well known work on intra-industry trade by Grubel and Lloyd (1975). These works were then followed by, what we know as today as, the theory of intra-industry trade. These models of intra-industry trade developed on the heels of the work on monopolistic competition and product differentiation by Dixit and Stiglitz (1977). They began with the work of Krugman (1979, 1980, 1981), Lancaster (1980), and Helpman (1981) who developed models of horizontally differentiated intra-industry trade with monopolistic competition—these models, and their derivatives, are summarized in Helpman and Krugman (1985). Models of horizontally differentiated intra-industry trade under oligopolistic competition, of the form of Eaton and Kierzkowski (1984), followed shortly after. Vertically differentiated intra-industry trade with perfect competition has been modelled by Caves (1981) using a version of the classical Heckscher-Ohlin international trade model as well as oligopolistic models of vertically differentiated intra-industry trade, such as Shaked and Sutton (1984).

Empirical investigations of these models are omnipresent. They cover issues dealing with the measurement of intra-industry trade and the estimation of its determinants based on the models named above. Generally focussing on determinants most of the models have in common, these studies investigate intra-industry trade between and within both developed and developing countries. This paper brings together the various measures and econometric studies on intra-industry trade into one place. A complete review of the literature is well beyond the scope of this paper, but it highlights the general trends and common features in measurement and econometric studies, as well as the advancements that have taken place over the development of this literature.

Measuring Intra-industry Trade

The purpose of this section is to discuss the methodological aspects of the measurement of intra-industry trade, rather than a documentary study of intra-industry trade applied to a country, or group of countries. Most of the references in this section, as well as the section below on the determinants of intra-industry trade, contain documentation of intra-industry trade levels for the countries under study that are too numerous to mention here in this review.

Usually, intra-industry trade is defined as the simultaneous export and import of goods in the same industry. But before we can discuss any measurement of intra-industry trade, we must decide what we are to measure. This is not a philosophical question, but a practical one as the history of empirical intra-industry trade has been mired by allegations of being a "statistical phenomenon" (Lipsey, 1976). The charge of being a statistical phenomenon is not an idle one. At the 3-digit SITC level of aggregation, canoes and 200,000 tonne tankers are in the same "ships and boats" industry; at the same level of aggregation, table model radios and airport flight control equipment are in the "telecommunications apparatus" industry (Lipsey, 1976). Also, Finger (1975) notes that trade overlap is not inconsistent with classical trade theory if empirical product groups do not correspond with the appropriate factor proportions groupings.

Industry/product categories have become sufficiently disaggregated to disregard these early claims of intra-industry trade being a statistical phenomenon. The Harmonized Tariff Schedule (HTS) has a 10-digit classification system with over 20,000 entries that not only separate canoes from 200,000 tonne tankers, but also from any other boat not designed to be used with motors or sails—a classification that is even distinct from a rowboat. However, aggregation should not be forgotten since it may still be an issue for other reasons. Nevertheless, we will now move on to the various measures.

7.4 The Balassa and Grubel-Lloyd Indices

Balassa (1966) proposed the first index of intra-industry trade that measured the degree of trade overlap—simultaneous import and export—of goods within an industry :

$$B_j \;=\; \frac{|X_i - M_i|}{(X_i + M_i)}. \qquad \ldots (1)$$

where $i \equiv$ commodity within industry j. This index, the ratio of net trade to gross trade, ranging from 0 to 1, with 0 representing "perfect" trade overlap, and therefore pure intra-industry trade, while 1 represents pure inter-industry trade. In order to calculate the degree of intra-industry trade for all industries (country level), Balassa took an unweighted average for each B_j :

$$B \;=\; \frac{1}{n} \sum B_j \qquad \ldots (2)$$

where $n \equiv$ number of industries. This can be generalized to be a weighted index :

$$B \;=\; \sum_j w_j B_j \qquad \ldots (3)$$

where $w_j \equiv$ industry j's share of total trade.

Though the essence of this index has remained intact to this day, an index that measured intra-industry trade that gave pure intra-industry trade a value of zero was not intuitively appealing. Grubel and Lloyd (1975) proposed an alternative index :

$$GL \;=\; \frac{(X_i + M_i) - |X_i - M_i|}{(X_i + M_i)} = 1 - \frac{|X_i - M_i|}{(X_i + M_i)} = 1 - B_j \qquad \ldots (4)$$

where $i \equiv$ commodity within industry j, that assigned pure intra-industry trade a value of 1 and pure inter-industry trade a value of 0. As with the Balassa Index, the Grubel-Lloyd Index has been calculated as an (un) weighted average to measure the degree of intra-industry trade at the country level.

This class of index has been criticized for suffering from categorical/sub-group aggregation issues. These issues have two basic forms that bias the index towards 1 : the grouping of two products in the same industry that should not be classified together, the canoe and tanker example above; and trade imbalance. The grouping of two, or more, categories together that should not be in the same industry is best explained using the following table :

Table 1 : Simple aggregation bias in the GL Index

Category	X_i	M_i	$|X_i - M_i|$	$(X_i + M_i)$	GL Index
3-Digit	150	160	10	310	0.968
Sub-Group 5-Digit	0	160	160	160	0.00
Sub-Group 5-Digit	150	0	150	150	0.00

Suppose we have one 3-digit "industry" that contains 2 sub-groups and each sub-group is independently engaged in (pure) inter-industry trade. We can see that the Grubel-Lloyd Index is zero for each of these sub-groups, so if we took an average, weighted or unweighted, of the two, the Grubel-Lloyd Index would still be zero. If, however, the import and export values are summed to form the 3-digit category, it appears that we have almost pure intra-industry trade with a Grubel-Lloyd Index of 0.968. Though this is an extreme example, it should be clear that aggregating across improper categories can lead to a misrepresentation of the degree of intra-industry trade.

The simple aggregation bias example above is a particular case of trade imbalance bias—trade imbalance, however, can occur when sub-groups are appropriately aggregated. This problem arises

when the net trade-gross trade ratio is characterized by opposite trade imbalances for the sub-groups (Greenaway and Milner, 1983). Suppose there are two commodities/sub-groups within an industry:

$$\frac{\left|X_i - M_i\right|}{(X_i + M_i)} = \frac{\left|(X_{1i} - M_{1i}) + (X_{2i} - M_{2i})\right|}{(X_{1i} + X_{2i} + M_{1i} + M_{2i})} \qquad \text{... (5)}$$

If the country in question is a net exporter (importer) in both sub-groups the weighting effect of the ratio is maintained, but if the country is a net exporter of one good and a net importer of the other good, the weighting effect is lost and the Grubel-Lloyd Index will take on a different value (Greenaway and Milner, 1983). This can be seen in the following table :

Table 2 : Trade imbalance bias in the GL Index

Category	X_i	M_i	$\|X_i - M_i\|$	$(X_i + M_i)$	GL Index
3-Digit	180	310	130	490	0.735
Sub-Group 5-Digit	80	160	80	240	0.667
Sub-Group 5-Digit	100	150	50	250	0.800
3-Digit	230	260	30	490	0.939
Sub-Group 5-Digit	80	160	80	240	0.667
Sub-Group 5-Digit	150	100	50	250	0.800

In the first category the country is a net importer in both sub-groups, but in the second category the country is a net importer in one good and a net exporter in the other. Since the Grubel-Lloyd Index does not recognize the direction of trade, the sub-group Grubel-Lloyd Indices are the same in both cases, but when the sub-groups are aggregated the Grubel-Lloyd Index for the second category is biased upward.

The index can be corrected by replacing the original net trade-gross trade ratio with the following net trade-gross trade ratio :

$$\frac{\sum_{i=1}^{n}\left|X_{ij} - M_{ij}\right|}{(X_j + M_j)} \qquad \text{... (6)}$$

where $i \equiv$ sub-group *i* within industry *j*. This adjustment removes the trade imbalance bias that results from countries being a net exporter in one sub-group of an industry and a net importer in another sub-group as well as the simple aggregation bias. We are left with the following index of intra-industry trade :

$$GL'_j = 1 - \frac{\sum_{i=1}^{n}\left|X_{ij} - M_{ij}\right|}{(X_j + M_j)} \qquad \text{... (7)}$$

Generally speaking, if a country is a net exporter/importer in both goods, GL = GL', but if a country is a net exporter in one good and a net importer in another, GL > GL' : $0 \le GL' \le GL \le 1$ (Greenaway and Milner, 1983). There was another adjustment suggested to the Grubel-Lloyd Index by Aquino (1978) in response to an imbalance in *overall* trade. Greenaway and Milner (1981) subsequently showed that the suggested adjustment is more likely to induce, rather than remove, distortions in the Grubel-Lloyd Index. Not surprisingly, this Aquino adjustment has fallen out of favour.

7.5 Measuring Marginal Intra-Industry Trade

Despite the ability to calculate the Grubel-Lloyd Index over time, it does not have desirable dynamic properties. An increase or decrease in the Grubel-Lloyd Index is not necessarily associated with corresponding increases or decreases in intra-industry trade. Caves (1981) and Hamilton and Kniest (1991) have noted that an equal/proportional increase in the exports and imports within an industry

from trade liberalization would raise the quantity of intra-industry trade, but it's proportion measured by the Grubel-Lloyd Index would remain the same. Suppose that trade liberalization doubled both imports and exports within a particular industry.

Table 3 : The dynamics of the GL Index

Category	X_i	M_i	$|X_i - M_i|$	$(X_i + M_i)$	GL Index
Pre-Liberalization	200	100	100	300	0.667
Pre-Trade Barrier	200	100	100	300	0.667
Post-Liberalization	400	200	200	600	0.667
Post-Trade Barrier	100	100	0	200	1.00

We can see from Table 3 that the net trade-gross trade ratio is multiplied and divided by the same scalar, two in this case, the value of the net trade-gross trade ratio, and hence the Grubel-Lloyd Index, remains the same :

$$\frac{|2X_i - 2M_i|}{(2X_i + 2M_i)} = \frac{2|X_i - M_i|}{2(X_i + M_i)} = \frac{|X_i - M_i|}{(X_i + M_i)}. \qquad \text{... (8)}$$

Also in Table 3, we see the possible effect of some trade barrier imposed. The exports of one country are decreased, and thus the quantity of intra-industry trade has fallen, yet this decrease in exports has put the two countries in perfect trade balance in this industry. The Grubel-Lloyd Index has actually risen from 0.667 to 1.00 even though intra-industry trade has decreased. This does not mean the Grubel-Lloyd Index is of no use when comparing trade over time, we must simply be cautious when interpreting change in the index.

Changes in intra-industry trade over time have significant effects on adjustment costs resulting from that change in trade — adjustment costs that have no doubt been taking place in recent years from the implementation of the FTA, NAFTA, and EU. In the first work on empirical intra-industry trade, Balassa (1966) noted that due to the presence of intra-industry trade, difficulties of adjustment have been overstated. Of particular importance to Canadian trade, if the FTA and/or the NAFTA brought about adjustment within the motor vehicle industry — manufacturing a different type of motor vehicle or switching to parts manufacturing — these adjustment costs would be much less than adjustment from the motor vehicle industry to another industry, such as textiles. Manufacturing a different type of motor vehicle, whether it be different on the quality or variety spectrum, would most likely entail similar production methods and employment practices such that any adjustment process would not be difficult. In fact, this is an adjustment that occurs quite regularly with the introduction of new automobile models. Even a switch from automotive manufacturing to automotive parts manufacturing would benefit from previous industry knowledge; automobile and automotive parts manufacturers would necessarily have knowledge of each others' markets since one supplies the other with an intermediate good(s). Production methods, as well as employment practices, would undergo much more change than the previous example, but not as much as a switch to the textile industry.

Due to the concern of measuring adjustment costs due to trade liberalization and because of the dynamic problem of the Grubel-Lloyd Index, a variant of the Grubel-Lloyd Index, called the Marginal Intra-Industry Trade Index, was developed by Hamilton and Kniest (1991) :

$$\text{M I I T} = \begin{cases} \dfrac{X_t - X_{t-n}}{M_t - M_{t-n}} & \text{for} \quad M_t - M_{t-n} > X_t - X_{t-n} > 0 \\[2mm] \dfrac{M_t - M_{t-n}}{X_t - X_{t-n}} & \text{for} \quad X_t - X_{t-n} > M_t - M_{t-n} > 0 \\[2mm] \text{undefined} & \text{for} \quad X_t < X_{t-n} \, \text{or} \, M_t < M_{t-n} \end{cases} \qquad \text{... (9)}$$

where n is the number of years between the two years of measurement. This index of marginal intra-industry trade captures the proportion of the increase in exports (imports) within an industry with a corresponding increase in imports (exports) within that same industry. Since this index will only measure new trade flows, by definition, it captures the relative importance of intra-industry trade generated by trade liberalization. As with the Grubel-Lloyd Index, the Marginal Intra-Industry Trade Index takes on values between 0 and 1, with 1 representing new trade that is pure intra-industry trade (Hamilton and Kniest, 1991).

We now have a representation of the dynamic nature of inter- and intra-industry trade for the purpose of evaluating adjustment costs over some time period. However, as with most first attempts, this index has complications. Greenaway, Hine, Milner and Elliott (1994) state that this index of marginal intra-industry trade that is undefined whenever ΔX or ΔM is negative ignores precisely what it is trying to measure. Using United Kingdom trade data, they find that no less than 32 percent of all 5-digit SITC categories are undefined by this index. Also, this measure indicates the importance of new intra-industry trade without any reference to the amount of new trade—a high index value may not be meaningful. There is also a problem of inflation causing an upward bias in this measure if the same quantity of exports (imports) now commands an inflated price. This will give the appearance of increased intra-industry trade that was really a nominal phenomenon; using real-valued trade data easily corrects for this bias (Greenaway *et al.*, 1994).

Greenaway *et al.* (1994) propose the following index, which differs from the Hamilton and Kniest (1991) index by representing intra-industry trade in values, rather than as a ratio :

$$M\,I\,I\,T' = [(X + M) - |X - M|]_t - [(X + M) - |X - M|]_{t-n} \qquad \dots (10)$$

$$= \Delta\,[(X + M) - |X - M|]. \qquad \dots (11)$$

As a consequence, this ratio is always defined and can easily be related to levels of new trade in order to assess the significance of this new trade. However, this measure suffers from the same trade imbalance bias discussed with the Grubel-Lloyd Index above, which was precisely the criticism held by Hamilton and Kniest (1991).

Brülhart (1994) suggests an index of marginal intra-industry trade that is always defined and does not suffer from the trade imbalance bias of previous indices :

$$M\,I\,I\,T'' = 1 - \frac{\left|(X_t - X_{t-n}) - (M_t - M_{t-n})\right|}{\left|X_t - X_{t-n}\right| + \left|M_t - M_{t-n}\right|} = 1 - \frac{|\Delta X - \Delta M|}{|\Delta X| + |\Delta M|}. \qquad \dots (12)$$

As with previous indices, this index takes on values between 0 and 1, with 1 representing pure marginal intra-industry trade. Like the Hamilton and Kniest Index, this index of marginal intra-industry trade captures the nature of the change in export and import flows, which is the desired property of such an index. In order to ensure this index is of economic significance, one only needs to take reference to the absolute (real) value of new trade.

Brülhart (1994) has also suggested an index of marginal intra-industry trade to capture industry performance; this index will allow for an investigation into the distribution of trade-induced gains (losses) between countries :

$$M\,I\,I\,T'' = \frac{\Delta X - \Delta M}{|\Delta X| + |\Delta M|}. \qquad \dots (13)$$

Unlike previous Grubel-Lloyd type indices, this index of marginal intra-industry trade ranges between –1 and 1. The closer M I I T''' is to zero, the higher is marginal intra-industry trade, whereas values

close to –1 and 1 represent higher marginal *inter*-industry trade : if $\Delta X > \Delta M$, $M\,I\,I\,T'''> 0$, and if $\Delta X < \Delta M$, $M\,I\,I\,T''' < 0$. So, values of $M\,I\,I\,T''' > 0$ indicate that exports are expanding at the expense of imports (strong domestic industry performance), conversely for $M\,I\,I\,T''' < 0$ (weak domestic industry performance). Unlike previous Grubel-Lloyd type indices, this index cannot have a (un) weighted average taken to assess marginal intra-industry trade at the country level; an average of – 1 and 1 is zero, which grossly misrepresents the type of trade.

One final issue with the measurement of marginal intra-industry trade comes from Thom and McDowell (1999). They claim that the Brülhart Index cannot distinguish between inter-industry trade and vertical intra-industry trade, and therefore, overestimates the costs of adjustment due to changes in trade composition—though the costs of adjustment for vertical intra-industry trade (quality differentiated goods) may be higher than horizontal intra-industry trade (variety differentiated goods), both will have lower adjustment costs than inter-industry trade. As we shall see in the next section, this is not entirely true. The Brülhart Index captures all intra-industry trade for which there is a simultaneous export and import of the same commodity classification; this includes horizontal intra-industry trade and vertical intra-industry trade that is defined by quality differentiation, but does not include vertical intra-industry trade along the lines of vertical integration within an industry— two, or more, distinct commodities traded between two countries, which are usually deemed as being within the same industry. The latter of the definitions of vertical intra-industry trade is usually not considered in empirical studies, and for good reason; grouping distinct commodity classifications together, although intuitively appealing at times since different sizes of automobiles have distinct commodity classifications, returns us to the difficulties of categorical aggregation discussed above. Commodity categories have become sufficiently disaggregated to avoid categorical aggregation issues and meaningfully disentangle vertical and horizontal product differentiation, but we must be careful not to take two steps backward from this one step forward.

7.6 Disentangling Horizontal and Vertical Intra-Industry Trade

Although the measurement of intra-industry trade as a whole has come a long way since Balassa (1966) first proposed a measure, we have seen above, for reasons of measuring adjustment costs, that there are reasons to disentangle horizontal and vertical intra-industry trade from each other. Also, as we will see below, vertical and horizontal intra-industry trade have different expectations with respect to parameter values of the determinants of intra-industry trade. We will now discuss how these two trade types have been disentangled in the literature.

Unfortunately, the Grubel-Lloyd Index, and its variants, gives us no explicit way to differentiate between one-way an two-way trade; the index tells us the degree of trade overlap, but doesn't tell us when we are dealing with two-way trade. If we are to take the definition of two-way trade literally, the simultaneous import and export of the same commodity classification, any commodity that has a Grubel-Lloyd Index greater than zero will be deemed two-way trade. More generally, we can consider trade within a commodity classification two-way trade when the value of the minority value flow of trade represented at least γ percent of the majority value flow of trade, which is the following condition :

$$\text{Two-way trade if :} \quad \frac{\text{Min}\left(X_{p,t}, M_{p,t}\right)}{\text{Max}\left(X_{p,t}, M_{p,t}\right)} > \gamma\% \qquad \text{... (14)}$$

where $p \equiv$ product and $t \equiv$ year. Below this level, the minority value flow would not be considered significant since it does not represent a structural feature of trade. This criterion can then be used to calculate an index of two-way trade :

$$\text{Share of Two-Way Trade} = \frac{\sum_i (X_i + M_i)}{\sum_j (X_j + M_j)} \qquad \text{... (15)}$$

where $i \equiv$ two-way traded goods and $j \equiv$ all traded goods.

Abd-el-Rahman (1991) pioneered this method in disentangling intra-industry trade. The index of two-way trade, however, had been proposed by Fontagnè, and Freudenberg (1997). Although the Grubel-Lloyd Index and the Two-Way Trade Index measure two different phenomenon — the Grubel-Lloyd Index measures the degree of trade overlap, while the two-way trade index considers all trade over the γ percent threshold to be two-way trade — when they are compared, they are quite similar. Fontagnè, and Freudenberg (1997), using regression analysis and a quadratic specification, found the fit between the two indices to be impressive : $R^2 = 0.97$. Given the longevity of the Grubel-Lloyd Index, this goodness of fit has provided some comfort to researchers.

Thus far, we have only differentiated between one- and two-way trade. We now must move to disentangle horizontal and vertical intra-industry trade. Within a given commodity classification that experiences two-way trade, products may or may not differ in their quality. In models of intra-industry trade, horizontal product differentiation is characterized by products with similar quality levels, with different attributes, while vertical differentiation is characterized by products with significantly different quality levels. Following Stiglitz (1987), empirical work that has disentangled intra-industry trade has assumed that prices represent quality, even under imperfect information. From this assumption, differences in the unit values (UV) or prices of these commodities can be assumed to represent these quality differences. Unit values have been defined for each commodity classification as the value of trade divided by the quantity traded, giving an average price of the goods traded in this category. Clearly, the more disaggregated the classification system, the better this method will be in capturing the price of the commodities. A classification system such as the 10-digit Harmonized Tariff Schedule with 20,000 commodity classifications will capture this well. The categories are so specific that different commodities will have different quantity measures : liters, kilograms, number, etc. while the SITC classification system is more general and uses tonnes as its quantity variable for all commodity categories.

Regardless of the level of disaggregation, horizontal product differentiation is defined as having the ratio of the export unit value to the import unit value falling within a range :

$$1 - \alpha \le \frac{UV^X}{UV^M} \le 1 + \alpha \qquad \text{... (16)}$$

where α is the threshold for the range. Vertical product differentiation is then defined as :

$$\frac{UV^X}{UV^M} > 1 + \alpha \quad \text{or} \quad \frac{UV^X}{UV^M} < 1 - \alpha \;. \qquad \text{... (17)}$$

Fontagnè, and Freudenberg (1997) have suggested a modified criteria that preserves the relative nature of the threshold :

$$\frac{1}{1 + \alpha} \le \frac{UV^X}{UV^M} \le 1 + \alpha \qquad \text{... (18)}$$

for horizontal product differentiation, and :

$$\frac{UV^X}{UV^M} > 1 + \alpha \quad \text{or} \quad \frac{UV^X}{UV^M} < \frac{1}{1 + \alpha} \qquad \text{... (19)}$$

for vertical product differentiation. For small values of α there is little difference between the two methods, but as α gets large the relative "distance" from the lower bound to unity becomes increasingly larger than the distance from unity to the upper bound.

We can see from Table 4 that as α increases the unit value ratio range becomes increasingly skewed to the lower bound in accounting for horizontal product differentiation. The modification suggested, and used, by Fontagnè, and Freudenberg (1997) corrects for this difficulty.

Table 4 : Differences in product quality measures

α	Relative Distance to Unity from Lower Bound $1 - \alpha$	Relative Distance to Unity from Lower Bound $1/(1 + \alpha)$
5 %	5%	5%
10%	11%	10%
15%	18%	15%
20%	25%	20%
25%	33%	25%

The two thresholds used for the distinction between vertical and horizontal product differentiation in the literature are 15 and 25 percent. The 15 percent threshold is generally used, and considered appropriate, when price differences reflect only differences in quality—the assumption of perfect information, such that a consumer will not purchase a similar, or lower, quality good at a higher price. However, in case of imperfect information the 15 percent threshold may be too narrow and the 25 percent threshold may be more appropriate. Alternatively, both thresholds could be used in order to evaluate the robustness of results (Greenaway, Hine, and Milner, 1995). Thus far, results of work on intra-industry trade have not been sensitive to the choice of the threshold.

The preceding criteria for trade overlap and product similarity lead to three different categories of trade :

1. two-way trade in similar, horizontally differentiated products (significant overlap and low unit value differences)

2. two-way trade in vertically differentiated products (significant overlap and high unit value differences)

3. one-way trade (no significant overlap).

With quality ranges of goods defined as up-market, middle-market, and down-market goods :

- up-market : unit value ratio $> 1 + \alpha$

- middle-market : $1/(1 + \alpha) \leq$ unit value ratio $\leq 1 + \alpha$

- down-market : unit value $< 1/(1 + \alpha)$,

one can investigate which price/quality segments of the market countries or industries lie, or move towards over time.

The share of two-way trade in horizontally differentiated products in industry j, the ratio of the value of two-way trade for which UV^X/UV^M falls within the horizontally differentiated products range,

$1/(1+\alpha) \leq \dfrac{UV^X}{UV^M} \leq 1+\alpha$, to the total value of trade in that industry is calculated :

$$TW\,H\,D_j \;=\; \frac{\sum_{p_i \in j} \sum_{HD} \left(X_{p,t} + M_{p,t}\right)}{\sum_{p_i \in j} \sum_{Z} \left(X_{p,t} + M_{p,t}\right)} \qquad \text{... (20)}$$

where TW $HD_j \equiv$ two-way horizontally differentiated trade share, $HD \equiv$ horizontally differentiated trade, $Z \equiv$ all trade types, $p_i \in j \equiv$ product i in industry j, and $t \equiv$ year. A similar formula is used in the calculation of the share of two-way trade in vertically differentiated products in industry j; that is, when

$$\frac{UV^X}{UV^M} < 1/(1+\alpha) \text{ or } \frac{UV^X}{UV^M} > 1 + \alpha :$$

$$TWVD_j = \frac{\sum_{p_i \in j} \sum VD\left(X_{p,t} + M_{p,t}\right)}{\sum_{p_i \in j} \sum Z\left(X_{p,t} + M_{p,t}\right)} \qquad \dots (21)$$

where $TWVD_j \equiv$ two-way horizontally differentiated trade share, $VD \equiv$ horizontally differentiated trade, $Z \equiv$ all trade types, $p_i \in j \equiv$ product i in industry j, and $t \equiv$ year. And of course, the share of one-way trade in industry j would be calculated as follows :

$$OWT_j = 1 - TWHD_j - TWVD_j \qquad \dots (22)$$

where $\qquad\qquad\qquad\qquad OWT_j \equiv$ one-way trade share.

Subsequent empirical work on the determinants of intra-industry trade by Greenaway, Hine, and Milner (1994, 1995) and Greenaway, Milner, and Elliott (1999) have used the initial threshold measure of product quality initiated by Abd-el-Rahman (1991) and a trade overlap value of γ = 0 percent, while Fontagnè, Freudenberg, and Pèridy (1997) have used the alternative threshold measure of product quality provided by Fontagnè, and Freudenberg (1997) and a trade overlap value of γ = 10 percent.

The Determinants of Intra-Industry Trade

In the previous section we discussed a number of ways that intra-industry trade has been measured. Though there may be some difficulties and issues with the measurement of intra-industry trade there has been a general consensus in the literature of what to measure. The same cannot be said for the determinants of intra-industry trade.

Since the first models of intra-industry trade, a great diversity of models in both horizontal and vertical intra-industry trade as well as alternative market structures such as monopolistic competition and oligopoly have come to pass. Some of these models have differing determinants/predictions while others have determinants that would prove difficult to discriminate between. Despite these difficulties, a multitude of empirical studies have sought to identify characteristics that are common to all, or most, of these models. These characteristics, of course, are subject to measurement error and are in large part proxy variables, which makes some of the measurement issues above seem insignificant. These characteristics have bee broadly classified as country-specific and industry specific determinants (Greenaway and Milner, 1989).

7.7 Intra-Industry Trade Theory and Developing Economics

As we have seen from our previous discussions that the first generation models in this new literature have been shown to be applicable for explaining trade between countries at similar levels of per capita income or development viz., North-North trade. Krugman argues that this exclusion of "poor nations" is due to "a bias in the research agenda". The trade analysis of LDCs requires taking account of increasing returns and imperfect competition. The "new theory' is essentially based on "arbitrary" specialization to realize economies of scale. The specific characteristics of LDCs require a different emphasis for adopting the theories to the developing world. Certain features of Southern economies are of relevance that bears on appropriate choice of models of trade. One pertinent point to note is that most of the theories are applicable to manufactures trade not the 'climatic' primary products. It is widely recognized that the simple H-O. theory is capable of explaining N-S trade in terms of the differences in factor endowment and mutual gains from this exchange are assured. The preference similarity theory, for example, is primarily applicable to N-N trade with tastes being determined by income level. The South having substantially lower income levels, on average, and having different tastes from the North would not participate in this form of trade. Marked income inequalities among the South economies imply that having skewed income distribution, the rich in these economies have incomes and hence, tastes similar to those in the North. South could gain from preference-similarity

trade with other Southern economies only if the South innovates and produces its own products. In these types of models, Southern focus is left ignored because the South has a comparative disadvantage in the production of these goods.

Where the South concentrates on production of differentiated consumer goods, it does it on a small scale and potential economies of scale remain partially, if not fully, unexploited. Stewart (1984) discusses this case in detail by pointing out that a lack of natural comparative advantage in their production at the primary stages when there are diseconomies of scale and no inefficiency (i.e., under employment of resources, etc.) compel most of them to produce under heavy protections. As Stewart recognizes, there is a trade off between efficiency and variety, unless trade can perform the role of permitting specialization, scale economies and variety to be exploited. As far as intra-industry specialization and exchange is concerned, it is solely a matter of more efficient deployment of resources within the South, and it should raise productivity and not direct resources. This is a form of trade creation. If countries in the South specialized on particular differentiated variety of final products and then exchanged them each country should be able to use its resources more efficiently, raising output without resource diversions. As the South moves into 'manufactured exports, as we will see in the subsequent chapters, the need for analyzing this kind of specialization and gains from exchange is looming large. The most general conclusion, following Stewart, is that while the theories explain N-N trade, S-S trade offer a potential way for the South to gain from the trade in products for countries with identical demand structure, in differentiated products and in intermediate goods. All of the models discussed in Section II postulate an equivalence between scale economy and IIT in the sense that for individual product varieties within an industry (e.g., car) having a common technology there are scale economies (internal to the firm), giving rise to IIT or 'non-comparative' advantage trade. This equivalence does not necessarily hold for the LDC because for small open economies the realization of it requires inter-industry trade. It is due to the fact that scale economies apply to large product lines and also a single scale efficient plant often exceeds the domestic market size of many LDCs. To take advantage, many industrial complexes will concentrate in a single country giving rise to IIT and specialization.

Krugman (1988) describes a situation (hypothetical) where scale economies in infrastructure (required for imports into agriculture) give the overall primary good a definite 'non-comparative advantage 'in trade pattern. In LDCs exports of manufacturers, there are elements of increasing returns as well as comparative advantage; hence, the phenomenon of IRS may be applicable to LDCs, even if primarily in the role of providing necessary infrastructure to get them to world or, in non-traded intermediate inputs to standard primary productive activity. The economic integration efforts in the developing world were always motivated more by the "swapping of production for import substituting industries' enjoying scale economies". Scale economies have mattered much in LDC trade policy. Since the market structure matters in international exchanges, it is necessary to analyze the imperfections in market characteristics in such 'regulation-prone' setting for LDCs. Firms and government interact in arriving at regulations and controls which have a spillover effect on a region's trade pattern. Small number of private agents and relatively heavy intrusion of government regulation make the traditional assumption of large, competitive markets, prima facie, less relevant for LDCs than the developed countries (DCs). For many LDCs these gains do not accrue to the state. Brander and Speneer (1981) type strategic export subsidies may seem a little far fetched for LDCs as their bargaining ability and credibility of threats is small. Even if the small country has some major share of the product, the government will face difficulty of becoming a credible first mover. Another interesting point to note is the fact that Krugman's 1984 paper "Import protection as export promotion": international competition in the presence of oligopoly and economies of scale' is a refurbished version of the 'Infant Industry Argument' for protecting an industry in a small LDCs. Krugman's model is based on IRS--internal to the firm so that firms in protected domestic market will move down their marginal cost curves and market shares will go up. The argument was originally pioneered by Hamilton (1791), List (1841), Mill (1909) and Bastable (1921). This is the model that countries like China, India, Japan and South Korea adopted and achieved international competitiveness.

Self-Assessment

1. **Choose the correct option**

    (*i*) Suppose Canada exports cars and imports furniture. This is an example of

    (*a*) Grubel-Lloyd trade.

    (*b*) Unbalanced trade, since cars are more valuable than furniture.

    (*c*) Intra industry trade.

    (*d*) Inter industry trade.

    (*ii*) Which of the following are true statements about intra-industry trade?

    (*a*) Low values on the Grubel-Lloyd index are associated with high levels of intra industry trade.

    (*b*) The majority of U.S. and European trade is intra industry trade.

    (*c*) Intra industry trade is especially common in agricultural sectors.

    (*d*) All of the above.

    (*iii*) If an industry is characterized by External Economies of Scale

    (*a*) Firm costs decline as the industry grows in size.

    (*b*) Firm costs decline as firms increase in size.

    (*c*) The market is likely to be served by an oligopoly of firms.

    (*d*) Monopolistic competition will follow.

    (*iv*) Consider a monopolistic market, such as the market for laptop computers. If a producer leaves the market,

    (*a*) Laptop prices will fall.

    (*b*) Laptop prices will rise and output per firm will increase.

    (*c*) Laptop prices will rise and output per firm will fall.

    (*d*) Output per firm will decrease as firms exploit market power.

    (*v*) Intra-industry trade benefits customers

    (*a*) In both countries, since prices fall and the number of varieties available to consumers increases.

    (*b*) In the larger country, since prices fall. Consumers in the smaller country are hurt by rising prices.

    (*c*) In both countries, since the effects of rising prices is offset by the benefit of an increased choice of varieties.

    (*d*) In the smaller country, since prices fall. Consumers in the larger country are hurt by rising prices.

7.8 Summary

*   Over the last decade or so there has been a reappraisal by some economists of the ability of traditional trade theory to explain relatively recent developments in the pattern of world trade. The 'new' views have largely focused on the economies of the DCs, newly industrialized countries (NICs), East Asian NICs and developing countries at higher stages of development.

*   These countries are rapidly dismantling their highly protectionist trade regimes often initiated by major multilateral lending agencies, including the IMF and the World Bank. There has been constant effort to determine from a theoretical standpoint whether or not there is anything in this "new" theory for small developing economies.

*   The new theories pay particular attention to two features of imperfect world markets, viz: (1) economies of scale; (2) external economies. These motives include the advantage of large scale production, technological advancement and innovation.

Notes
- As well, economic theorizing has improved, such that there is now a better understanding of industrial structure and the nature of oligopolistic markets. Herein lies the relevance of 'new view' for small open economies.
- Firms in these countries produce, nowadays, new high technology products of services for the export market, manufactures on a large scale by using high technology processes. This has got a strong empirical basis. Considerable dynamic benefits of the learning-by-doing type or 'technological invitation' variety may be expected to evolve through different ex ante policies and also through the adoption of development strategies.
- Although empirical work incorporating more recent theoretical approaches is not lagging behind, the present analysis shows an evolution of new trade patterns for the developing economies participating and integrating themselves with the world.

7.9 Key-Words

1. Intra-industry trade : Intra-industry trade refers to the exchange of similar products belonging to the same industry. The term is usually applied to international trade, where the same types of goods or services are both imported and exported.

2. International trade : It is the exchange of capital, goods, and services across international borders or territories. In most countries, such trade represents a significant share of gross domestic product (GDP). While international trade has been present throughout much of history, social, and political importance has been on the rise in recent centuries.

7.10. Review Questions

1. What are the causes of emergence of intra industry trade?
2. Discuss the measurement of intra industry trade and its impact on developing economics.

Answers: Self-Assessment

1. (*i*) (*d*)   (*ii*) (*b*)   (*iii*) (*a*)   (*iv*) (*c*)   (*v*) (*a*)

7.11 Further Readings

1. Altonji, Joseph, and David Card. 1991. "The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives." In John Abowd and Richard Freeman, eds, Immigration, Trade, and the Labor Market, (Chicago, IL: University of Chicago Press), 201-234.
2. Bernard, Andrew B. and J. Bradford Jensen. 1999. "Understanding Increasing and Decreasing Wage Inequality." In Robert C. Feenstra, ed., The Impact of International Trade on Wages, Chicago: University of Chicago Press.
3. Bernstein, Jeffrey R. and David E. Weinstein. 1998. "Do Endowments Predict the Location of Production? Evidence from National and International Data." NBER Working Paper No. 6815.
4. Blackorby, Charles, William Schworm, and Anthony Venables. 1993. "Necessary and Sufficient Conditions for Factor Price-Equalization." Review of Economic Studies 60: 413-434.

Unit 8 : Tariff, Quotas and Non-tariff Barriers : Definitions and Types

---

CONTENTS

Objectives

Introduction

8.1 Definition

8.2 Types of Tariffs

8.3 Quotas

8.4 Non-Tariff Barriers

8.5 Summary

8.6 Key-Words

8.7 Review Questions

8.8 Further Readings

Objectives

After reading this Unit students will be able to:

• Discuss the Definition and Types of Tariff and Non-tariff.

• Explain the Types of Quotas.

Introduction

A tax imposed on imported goods and services. Tariffs are used to restrict trade, as they increase the price of imported goods and services, making them more expensive to consumers. They are one of several tools available to shape trade policy. Governments may impose tariffs to raise revenue or to protect domestic industries from foreign competition, since consumers will generally purchase cheaper foreign produced goods. Tariffs can lead to less efficient domestic industries, and can lead to trade wars as exporting countries reciprocate with their own tariffs on imported goods. Organizations such as the WTO exist to combat the use of egregious tariffs.

Non-tariff barriers are another way for an economy to control the amount of trade that it conducts with another economy, either for selfish or altruistic purposes. Any barrier to trade will create an economic loss, as it does not allow markets to function properly. The lost revenues resulting from the barrier to trade can be called an economic loss.

Quota, in international trade, government-imposed limit on the quantity, or in exceptional cases the value, of the goods or services that may be exported or imported over a specified period of time. Quotas are more effective in restricting trade than tariffs, particularly if domestic demand for a commodity is not sensitive to increases in price. Because the effects of quotas cannot be offset by depreciation of the foreign currency or by an export subsidy, quotas may be more disturbing to the international trade mechanism than tariffs. Applied selectively to various countries, quotas can also be a coercive economic weapon.

Tariff quotas may be distinguished from import quotas. A tariff quota permits the import of a certain quantity of a commodity duty-free or at a lower duty rate, while quantities exceeding the quota are subject to a higher duty rate. An import quota, on the other hand, restricts imports absolutely.

If the quantity imported under a quota is less than would be imported in the absence of a quota, the domestic price of the commodity in question may rise. Unless the government maintains some system of licensing importers in order to capture as revenue the difference between the higher domestic price and the foreign price, the importing of such commodities can prove a lucrative source of private profit.

Quantitative trade restrictions were first imposed on a large scale during and immediately after World War I. During the 1920s quotas were progressively abolished and replaced by tariffs. The next great wave of quota protection came during the Great Depression in the early 1930s, with France leading the European countries in introducing a comprehensive quota system in 1931. After World War II, the western European countries began a gradual dismantling of quantitative import restrictions, but the United States tended to make more use of them.

8.1 Definition

Tariffs, which are taxes on imports of commodities into a country or region, are among the oldest forms of government intervention in economic activity. They are implemented for two clear economic purposes. First, they provide revenue for the government. Second, they improve economic returns to firms and suppliers of resources to domestic industry that face competition from foreign imports.

Tariffs are widely used to protect domestic producers' incomes from foreign competition. This protection comes at an economic cost to domestic consumers who pay higher prices for import-competing goods, and to the economy as a whole through the inefficient allocation of resources to the import competing domestic industry. Therefore, since 1948, when average tariffs on manufactured goods exceeded 30 percent in most developed economies, those economies have sought to reduce tariffs on manufactured goods through several rounds of negotiations under the General Agreement on Tariffs Trade (GATT). Only in the most recent Uruguay Round of negotiations were trade and tariff restrictions in agriculture addressed. In the past, and even under GATT, tariffs levied on some agricultural commodities by some countries have been very large. When coupled with other barriers to trade they have often constituted formidable barriers to market access from foreign producers. In fact, tariffs that are set high enough can block all trade and act just like import bans.

8.2 Types of Tariffs

Tariffs can be expressed in absolute or in relative terms, they may be discriminatory or non-discriminatory, they can be imposed on imports or on exports, and they may be prompted by considerations of revenue or protection to domestic industries.

They are expressed in absolute terms of dollars and cents per unit *e.g.* per tonne, or per pound of weight of the imported (or exported) quantity. Specific duties can be levied on goods like wheat or rice or sugar, but they can not be adopted for all goods and services especially in the case of valuable goods. For instance, specific tariffs can not be levied in the case of, say, diamonds, modern art paintings, transistors, television sets, etc. We cannot estimate duties on the articles by weighing them physically. The duty has to be estimated on the basis of the value of these products, rather than their physical weight. When, therefore, duty is levied on the basis of the value of the product measured by their money price, we have what are called as *ad valorem* tariffs *(ad valorem* is a Latin word which means "on the value"). Ad valorem tariff is a percentage tax.

Did u know?   *Specific tariffs* are those which are assessed on the basis of the physical weight of the product which is imported or exported.

There may also be a *compound tariff* which combines a specific duty with an *ad valorem* duty. The distinction between the two types of tariffs is of some significance. As the prices of imported goods rise, the *ad valorem* tariff based on a given fixed percentage brings greater revenue to the tariff imposing country, whereas specific tariffs lack such revenue elasticity with respect to import price changes.

A discriminatory tariff calls for different rates of duties depending on the country of origin or the destination of the product. For example, a country can impose higher rates of duty on goods coming into the country from, say, Australia and lower rates of duty on goods coming from Thailand. *A non-discriminatory tariff,* on the other hand, imposes a uniform rate of duty regardless of their source of origin. Tariffs are said to be *single column* when they are non-discriminatory, and *double column* when they are discriminatory.

Revenue tariffs are those which are imposed primarily to produce revenue for the government. With the introduction and expansion of income taxes and other direct taxes, the importance of tariffs as a source of revenue has considerably gone down, especially in developed countries. The less developed countries, however, still rely on tariffs as a substantial source of government revenues. In Malaysia, for example, the import and export duties together amount for well over 40 per cent of the total government revenues. In commercial policy, it is the *protective tariffs* that dominates the scene. When the tariffs are imposed primarily to protect the domestic industries from foreign competition, the country is said to have protective tariffs. The motive, here, is not revenue but protection of the domestic economy from foreign competition.

It is not possible to classify revenue tariffs and protective tariffs and put them into watertight compartments, because tariffs imposed for revenue will produce some protective effects and the protective tariffs yield some revenue as well. The difference between the two is basically with regard to their primary motive as such.

Finally, there are *retaliatory tariffs* and *countervailing tariffs.* When Country A imposes (or increases) duties against the products from Country B, it is possible that country B will retaliate and levy duties on goods imported from Country A. Country B's tariffs are then described as retaliatory tariffs. Their motive is neither to raise revenue nor to accord protection to domestic industries, but to act in retaliation. Tariffs are said to be countervailing when a country imposes (or increases) import duties with a view to offset export subsidy in the country of origin. For example, if Country B, the government of country B may think that country A is subsidizing its export to Country B, the government of country B may think that country A's products entering into Country B are enjoying "unfair advantage" over the country's domestic import replacement products, then Country B is justified in imposing countervailing duties on the products imported from Country A. The countervailing duties are primarily aimed at offsetting such an unfair advantage given by export subsidies in foreign countries.

Import duties have received most of the analytical and policy attention and they are far more widespread than the export duties. In some countries like the United States, export duties are prohibited by law. In other countries where export duties are in vogue, their sole aim appears to be revenue collection. Import duties, however, are motivated by two other considerations, besides revenue *viz.* considerations of protection and balance of payments adjustment. In what follows, therefore, we will concentrate our tariff study only with reference to the import tariffs.

8.3 Quotas

A quota is simply a maximum limitation, specified in either value or physical units, on imports of a product for a given period. It is enforced through licenses issued to either importers or exporters and may be applied to imports from specific countries or from all foreign countries generally. Two examples illustrate these different characteristics. The United States imposes a general quota on dried milk imports; licenses are granted to certain U.S. trading companies, who are allowed to import a maximum quantity of dried milk based on their previous imports. In a different situation U.S. sugar imports are limited by a quota that specifies the shares of individual countries; the right to sell sugar to the United States is given directly to the governments of these countries.

Import Quotas

A quota is a direct limitation of the physical quantity of exports and imports permitted in the country. We will only discuss import quotas as they are more common than the export quotas. The effects of quotas are similar to those of tariffs, but there are also substantive differences between the two which are worth examining.

Import quotas can be of different types. First of all, there can be *unilateral quotas* which can be *global* or *allocated.* In case of global quotas, quota is fixed in global terms and any foreign country can supply the amount fixed by quota at the lowest price. The quota imposing country gets the price advantage and the goal quotas are non-discriminatory. The allocated quotas, however, involve allocating and distributing the quota among different foreign countries on the basis of a pre-determined criterion. The allocated quotas are discriminatory; and the quota imposing country does not necessarily get the price advantage of the global quota system. Then there can be *bilateral quotas* as opposed to unilateral quotas. Bilateral quotas imply mutual agreement between countries through negotiation. The advantage of these quotas is that they do not provoke retaliation. Finally, there are what are called as mixing or indirect quotas, in whose case the domestic producers are asked to use a fixed proportion of imported and local materials used in producing their products. The quota is fixed not in absolute terms but in percentage terms.

Let us now come to the question of import quotas *versus* import duties and see their differences and similarities. Until the appearance of new contributions in the late 1960s and early 1970s, particularly made by Bhagwati and Corden, it was thought that quotas and duties were hardly distinguishable by their effects. The more recent studies have, however, shown there is no strict equivalence between the two.

There are certain similarities between the quotas and the duties

(a) they both have the same objectives, *viz.* to reduce the level of imports so as to protect domestic industries, correct balance of payments deficits, expand domestic employment and economic activities. While their objectives are the same, their methods are different. A tariff acts directly on the price and indirectly on the quantity of the imported goods. A quota acts directly on the quantity of imports and has an indirect effect on the price of the imported goods.

(b) A tariff of a certain height cuts imports to a certain quantity—it has, therefore, a quota equivalent. A quota would limit imports to a certain quantity and therefore, raises the import price—it has, thus, a tariff equivalent, and

(c) Since both quotas and tariffs raise the import price and reduce the import quantity they produce similar effects on consumption, production, trade balance, terms of trade, national income, redistribution, factor movements, economic growth and economic welfare.

The important differences between quotas and tariffs would be as follows :

(a) Tariffs bring revenues to the government whereas quotas do not. This raises certain redistribution and welfare issues. Under tariffs, a part of consumers' loss goes to the government by way of tariff revenue. Under the quota system the extent of consumers' loss will be more or less the same, but to whom that surplus goes is ambiguous. It could go to the government if the government charged a fee for selling import licences. In that case, the import licence fee, will be an equivalent of an import tariff. If, however, the quotas are distributed freely to the importers without charging a fee, then importer's surplus will increase at the cost of government revenue or consumer's welfare loss. Furthermore, tariff revenue can be used for social expenditure; but the quota profits, going to the importers, may not contribute to net social welfare.

(b) Distribution of import licences (associated with quotas) may give rise to corruption and bribery on the part of government officials. Import tariffs do not create such evils of government corruption, political favoritism, nepotism, etc.

(c) Quotas could be more effective than tariffs particularly when the domestic demand and supply curves for the import good are inelastic. See Figure 8.1.

If, as in the Figure 8.1, the demand and supply curves are inelastic, tariffs will fail to reduce import quantities. From the free trade price P_p the tariffs have raised the price to a substantially higher level, P_t, and yet the quantity imported has not been cut down by tariffs. Tariffs have, of course, resulted in substantial government revenues, but they have failed to generate protective

effects. Import quotas, if called upon, will certainly do a job of restricting imports to any desirable size and thereby create all the necessary protective effects. In that sense quotas succeed where tariffs may fail.
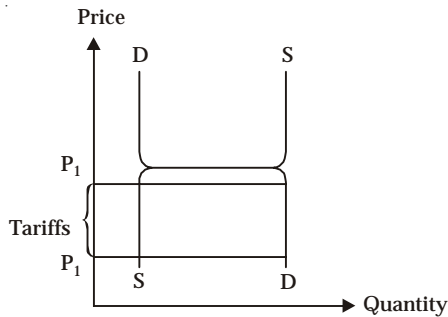
Figure 8.1 : Inelastic Demand and Supply Conditions Affecting Tariffs.

(d) The terms of trade effects of tariffs are determinate or predictable, but those resulting from quotas are indeterminate or unpredictable. See the Figure 8.2 A-B.

Tariff imposition, in Figure 8.1 A, shifts the tariff-imposing country's offer curve from OH to OH_1 which would result in an improvement in the country's term of trade. This is shown by the steeper, OT_1 terms of trade line (after tariffs) as compared with the free trade terms of trade line OT_0. In Figure 8.2 B, we show the result of quota imposition on the country's terms of trade. The quota-imposing country's offer curve is OH and the foreign country's offer curve is OF. Free trade equilibrium is at Q where the terms of trade are shown by the line OT_0. At trade equilibrium point Q, the country's imports are OM_0 (and exports are M_0Q); and the country may impose a quota on the imports such that the imports are reduced by the quota equal to the size of OM_1.
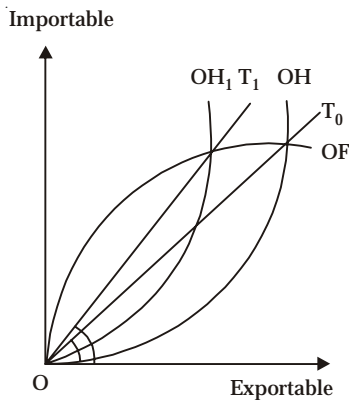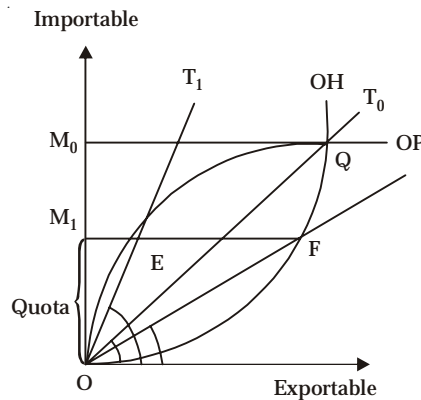
Figure 8.2 A   **Figure 8.2 B**

Diagram 8 A-B : Terms of Trade Effect of Tariffs and Terms of Trade Effect of Quotas.

The question that now arises is what would be the country's terms of trade as a result of fixing the import quota at OM_1 level. Now, draw a line from M_1 horizontally straight, and this line cuts foreign country's offer curve at E and the home country's offer curve at F. If, after the quota imposition, trade equilibrium takes place at E, then the quota-imposing country's (or home country's) terms of trade improve, because OT_1 is steeper than OT_0. On the other hand, if the new trade equilibrium takes place at F, then the country's terms of trade would worsen as indicated by the line OT_2 which is less steep compared to OT_0. This would make the terms of trade effects of quotas somewhat unpredictable, whereas the effects of tariffs on the country's terms of trade are somewhat predictable.

This argument, however, can not be stretched too far because, as we have seen earlier, even the effects of tariffs on the country's terms of trade movement are also not very clear-cut.

(e) There are several reasons why domestic producers, importers and even governments, may prefer quotas to tariffs. From the government's point of view, quotas are easy to impose, remove, change and administer. Secondly, international attitude is more permissive on quotas than on tariffs. Thirdly, high tariff levels necessary to achieve significant import reductions could seriously damage the public image of the protected industry. For this reason, the domestic producers may prefer "invisible' quota protection to the 'Visible' tariff protection. Fourthly, as the total demand for the product grows, a given quota fixed in absolute terms would be more protective to the domestic producers. A tariff, on the other hand, would not prevent foreign imports gaining their share of the expanded market in the protected sector of the economy. Fifthly, it may be argued that quotas allow at least some imports into the country, while tariffs may be too prohibitive to allow any imports at all. This may have relevance from the standpoint of consumers of the imported product. Finally, the possibility of capturing good profits from the quota would be a strong reason for the importers to press for quota method rather than the tariff method.

In conclusion, both quotas and tariffs provoke retaliation by the rest of the world. Both are anti-trade, and both tend to reduce economic welfare and allocative efficiency. And it is generally believed that quotas are more obnoxious than tariffs, because quotas carry additional costs, apart from those that are carried by the tariffs.

Effectiverate of Protection

Until recently, the degree of protection enjoyed by the domestic industry was thought to be a relatively simple matter. A ten per cent tariff on a finished imported product was said to give a ten per cent protection to the domestically produced import-replacement product. The degree of protection was, thus, measured by the rate of tariff duty imposed on the imported finished product. In the 1960s, however, a new concept was developed to measure the degree of protection which an industry enjoyed in a country. This is the concept of *effective* rate of protection, as opposed to the *nominal* rate of protection. The effective protection.

8.3 Definition and Types of Non-Tariff Barriers

Non-tariff barriers to trade (NTBs) are trade barriers that restrict imports but are not in the usual form of a tariff. Some common examples of NTB's are anti-dumping measures and countervailing duties, which, although called non-tariff barriers, have the effect of tariffs once they are enacted. Their use has risen sharply after the WTO rules led to a very significant reduction in tariff use. Some non-tariff trade barriers are expressly permitted in very limited circumstances, when they are deemed necessary to protect health, safety, sanitation, or depletable natural resources. In other forms, they are criticized as a means to evade free trade rules such as those of the World Trade Organization (WTO), the European Union (EU), or North American Free Trade Agreement (NAFTA) that restrict the use of tariffs.

Some of non-tariff barriers are not directly related to foreign economic regulations but nevertheless have a significant impact on foreign-economic activity and foreign trade between countries. Trade between countries is referred to trade in goods, services and factors of production. Non-tariff barriers to trade include import quotas, special licenses, unreasonable standards for the quality of goods, bureaucratic delays at customs, export restrictions, limiting the activities of state trading, export subsidies, countervailing duties, technical barriers to trade, sanitary and phyto-sanitary measures, rules of origin, etc. Sometimes in this list they include macroeconomic measures affecting trade.

Types of Non-Tariff Barriers

There are several different variants of division of non-tariff barriers. Some scholars divide between internal taxes, administrative barriers, health and sanitary regulations and government procurement policies. Others divide non-tariff barriers into more categories such as specific limitations on trade,

customs and administrative entry procedures, standards, government participation in trade, charges on import, and other categories.

The first category includes methods to directly import restrictions for protection of certain sectors of national industries : licensing and allocation of import quotas, antidumping and countervailing duties, import deposits, so-called voluntary export restraints, countervailing duties, the system of minimum import prices, etc. Under second category follow methods that are not directly aimed at restricting foreign trade and more related to the administrative bureaucracy, whose actions, however, restrict trade, for example : customs procedures, technical standards and norms, sanitary and veterinary standards, requirements for labeling and packaging, bottling, etc. The third category consists of methods that are not directly aimed at restricting the import or promoting the export, but the effects of which often lead to this result.

The non-tariff barriers can include wide variety of restrictions to trade. Here are some example of the popular NTBs.

Licenses

The most common instruments of direct regulation of imports (and sometimes export) are licenses and quotas. Almost all industrialized countries apply these non-tariff methods. The license system requires that a state (through specially authorized office) issues permits for foreign trade transactions of import and export commodities included in the lists of licensed merchandises. Product licensing can take many forms and procedures. The main types of licenses are general license that permits unrestricted importation or exportation of goods included in the lists for a certain period of time; and one-time license for a certain product importer (exporter) to import (or export). One-time license indicates a quantity of goods, its cost, its country of origin (or destination), and in some cases also customs point through which import (or export) of goods should be carried out. The use of licensing systems as an instrument for foreign trade regulation is based on a number of international level standards agreements. In particular, these agreements include some provisions of the General Agreement on Tariffs and Trade and the Agreement on Import Licensing Procedures, concluded under the GATT (GATT).

Licensing of foreign trade is closely related to quantitative restrictions—quotas - on imports and exports of certain goods. A quota is a limitation in value or in physical terms, imposed on import and export of certain goods for a certain period of time. This category includes global quotas in respect to specific countries, seasonal quotas, and so-called "voluntary" export restraints. Quantitative controls on foreign trade transactions carried out through one-time license.

Quantitative restriction on imports and exports is a direct administrative form of government regulation of foreign trade. Licenses and quotas limit the independence of enterprises with a regard to entering foreign markets, narrowing the range of countries, which may be entered into transaction for certain commodities, regulate the number and range of goods permitted for import and export. However, the system of licensing and quota imports and exports, establishing firm control over foreign trade in certain goods, in many cases turns out to be more flexible and effective than economic instruments of foreign trade regulation. This can be explained by the fact, that licensing and quota systems are an important instrument of trade regulation of the vast majority of the world.

The consequence of this trade barrier is normally reflected in the consumers' loss because of higher prices and limited selection of goods as well as in the companies that employ the imported materials in the production process, increasing their costs. An import quota can be unilateral, levied by the country without negotiations with exporting country, and bilateral or multilateral, when it is imposed after negotiations and agreement with exporting country. An export quota is a restricted amount of goods that can leave the country. There are different reasons for imposing of export quota by the country, which can be the guarantee of the supply of the products that are in shortage in the domestic market, manipulation of the prices on the international level, and the control of goods strategically important for the country. In some cases, the importing countries request exporting countries to impose voluntary export restraints.

Agreement on a "voluntary" export restraint

In the past decade, a widespread practice of concluding agreements on the "voluntary" export restrictions and the establishment of import minimum prices imposed by leading Western nations upon weaker in economical or political sense exporters. The specifics of these types of restrictions is the establishment of unconventional techniques when the trade barriers of importing country, are introduced at the border of the exporting and not importing country. Thus, the agreement on "voluntary" export restraints is imposed on the exporter under the threat of sanctions to limit the export of certain goods in the importing country. Similarly, the establishment of minimum import prices should be strictly observed by the exporting firms in contracts with the importers of the country that has set such prices. In the case of reduction of export prices below the minimum level, the importing country imposes anti-dumping duty, which could lead to withdrawal from the market. "Voluntary" export agreements affect trade in textiles, footwear, dairy products, consumer electronics, cars, machine tools, etc.

Problems arise when the quotas are distributed between countries because it is necessary to ensure that products from one country are not diverted in violation of quotas set out in second country. Import quotas are not necessarily designed to protect domestic producers. For example, Japan, maintains quotas on many agricultural products it does not produce. Quotas on imports is a leverage when negotiating the sales of Japanese exports, as well as avoiding excessive dependence on any other country in respect of necessary food, supplies of which may decrease in case of bad weather or political conditions.

Export quotas can be set in order to provide domestic consumers with sufficient stocks of goods at low prices, to prevent the depletion of natural resources, as well as to increase export prices by restricting supply to foreign markets. Such restrictions (through agreements on various types of goods) allow producing countries to use quotas for such commodities as coffee and oil; as the result, prices for these products increased in importing countries.

Notes

Embargo is a specific type of quotas prohibiting the trade. As well as quotas, embargoes may be imposed on imports or exports of particular goods, regardless of destination, in respect of certain goods supplied to specific countries, or in respect of all goods shipped to certain countries.

Although the embargo is usually introduced for political purposes, the consequences, in essence, could be economic.

Standards

Standards take a special place among non-tariff barriers. Countries usually impose standards on classification, labeling and testing of products in order to be able to sell domestic products, but also to block sales of products of foreign manufacture. These standards are sometimes entered under the pretext of protecting the safety and health of local populations.

Administrative and bureaucratic delays at the entrance

Among the methods of non-tariff regulation should be mentioned administrative and bureaucratic delays at the entrance, which increase uncertainty and the cost of maintaining inventory.

Import deposits

Another example of foreign trade regulations is import deposits. Import deposits is a form of deposit, which the importer must pay the bank for a definite period of time (non-interest bearing deposit) in an amount equal to all or part of the cost of imported goods.

At the national level, administrative regulation of capital movements is carried out mainly within a framework of bilateral agreements, which include a clear definition of the legal regime, the procedure

for the admission of investments and investors. It is determined by mode (fair and equitable, national, most-favored-nation), order of nationalization and compensation, transfer profits and capital repatriation and dispute resolution.

Foreign exchange restrictions and foreign exchange controls

Foreign exchange restrictions and foreign exchange controls occupy a special place among the non-tariff regulatory instruments of foreign economic activity. Foreign exchange restrictions constitute the regulation of transactions of residents and nonresidents with currency and other currency values. Also an important part of the mechanism of control of foreign economic activity is the establishment of the national currency against foreign currencies.

The Transition from Tariffs to Non-tariff Barriers

One of the reasons why industrialized countries have moved from tariffs to NTBs is the fact that developed countries have sources of income other than tariffs. Historically, in the formation of nation-states, governments had to get funding. They received it through the introduction of tariffs. This explains the fact that most developing countries still rely on tariffs as a way to finance their spending. Developed countries can afford not to depend on tariffs, at the same time developing NTBs as a possible way of international trade regulation. The second reason for the transition to NTBs is that these tariffs can be used to support weak industries or compensation of industries, which have been affected negatively by the reduction of tariffs. The third reason for the popularity of NTBs is the ability of interest groups to influence the process in the absence of opportunities to obtain government support for the tariffs.

Non-tariff barriers today

With the exception of export subsidies and quotas, NTBs are most similar to the tariffs. Tariffs for goods production were reduced during the eight rounds of negotiations in the WTO and the General Agreement on Tariffs and Trade (GATT). After lowering of tariffs, the principle of protectionism demanded the introduction of new NTBs such as technical barriers to trade (TBT). According to statements made at United Nations Conference on Trade and Development (UNCTAD, 2005), the use of NTBs, based on the amount and control of price levels has decreased significantly from 45% in 1994 to 15% in 2004, while use of other NTBs increased from 55% in 1994 to 85% in 2004.

Increasing consumer demand for safe and environment friendly products also have had their impact on increasing popularity of TBT. Many NTBs are governed by WTO agreements, which originated in the Uruguay Round (the TBT Agreement, SPS Measures Agreement, the Agreement on Textiles and Clothing), as well as GATT articles. NTBs in the field of services have become as important as in the field of usual trade.

Most of the NTB can be defined as protectionist measures, unless they are related to difficulties in the market, such as externalities and information asymmetries between consumers and producers of goods. An example of this is safety standards and labeling requirements.

The need to protect sensitive to import industries, as well as a wide range of trade restrictions, available to the governments of industrialized countries, forcing them to resort to use the NTB, and putting serious obstacles to international trade and world economic growth. Thus, NTBs can be referred as a new of protection which has replaced tariffs as an old form of protection.

Self-Assessment

1. **Choose the correct options:**

    (*i*) Protecting domestic producers against import competition:

        (*a*) Helps those producers.

        (*b*) Helps domestic consumers of the product.

        (*c*) Probably helps the importing nation as a whole.

        (*d*) All of the above.

(*ii*) A nontariff barrier operates by:

    (*a*) Limiting the quantity of imports.

    (*b*) Increasing the cost of getting imports to market.

    (*c*) Creating uncertainty about the conditions under which imports will be permitted.

    (*d*) All of the above.

(*iii*) One of the reasons that protectionists and government officials may favor using a quota instead of a tariff is:

    (*a*) Quotas generate more revenue for the government than do tariffs.

    (*b*) A quota ensures that the quantity of imports is strictly limited.

    (*c*) Quotas create less market distortions than do tariffs.

    (*d*) Quotas give less power to politicians than do tariffs.

(*iv*) A quota:

    (*a*) Causes domestic prices to fall.

    (*b*) Causes world prices to rise.

    (*c*) Restricts the quantity of a good that can be imported.

    (*d*) Is always more efficient than a tariff.

(*v*) In the case of a small country, a quota and a tariff are (almost) identical if:

    (*a*) The government allocates licenses for free to importers using a rule or process that involves (almost) no resource cost.

    (*b*) The government auctions off licenses to the highest bidder.

    (*c*) The government allocates licenses to importers through application and selection procedures that require the use of substantial resources.

    (*d*) The government allocates import licenses directly to the public using a free lottery system.

(*vi*) Which of the following is a means of allocating import licenses by assigning the licenses without competition, applications, or negotiation?

    (*a*) Fixed favoritism.

    (*b*) Resource-using application procedures.

    (*c*) Import-license auctions.

    (*d*) Domestic content requirements.

(*vii*) Which of the following requires that an import distributor buy a certain percentage of the product locally?

    (*a*) An import quota.

    (*b*) A mixing requirement.

    (*c*) A voluntary export restraint.

    (*d*) A domestic content requirement.

8.4 Summary

- Tariffs, which are taxes on imports of commodities into a country or region, are among the oldest forms of government intervention in economic activity.

- Tariffs are widely used to protect domestic producers' incomes from foreign competition. This protection comes at an economic cost to domestic consumers who pay higher prices for import-competing goods, and to the economy as a whole through the inefficient allocation of resources to the import competing domestic industry.

- Tariffs can be expressed in absolute or in relative terms, they may be discriminatory or non-discriminatory, they can be imposed on imports or on exports, and they may be prompted by considerations of revenue or protection to domestic industries.

- There may also be a *compound tariff* which combines a specific duty with an *ad valorem* duty. The distinction between the two types of tariffs is of some significance. As the prices of imported goods rise, the *ad valorem* tariff based on a given fixed percentage brings greater revenue to the tariff imposing country, whereas specific tariffs lack such revenue elasticity with respect to import price changes.

- *A discriminatory tariff* calls for different rates of duties depending on the country of origin or the destination of the product. For example, a country can impose higher rates of duty on goods coming into the country from, say, Australia and lower rates of duty on goods coming from Thailand. *A non-discriminatory tariff,* on the other hand, imposes a uniform rate of duty regardless of their source of origin.

- A quota is simply a maximum limitation, specified in either value or physical units, on imports of a product for a given period. It is enforced through licenses issued to either importers or exporters and may be applied to imports from specific countries or from all foreign countries generally. Two examples illustrate these different characteristics.

- There are several different variants of division of non-tariff barriers. Some scholars divide between internal taxes, administrative barriers, health and sanitary regulations and government procurement policies.

8.5 Key-Words

1. Tariff	:	A tax imposed on imported goods and services. Tariffs are used to restrict trade, as they increase the price of imported goods and services, making them more expensive to consumers. They are one of several tools available to shape trade policy.
2. Non-terrif Barriers	:	A form of restrictive trade where barriers to trade are set up and take a form other than a tariff. Nontariff barriers include quotas, levies, embargoes, sanctions and other restrictions, and are frequently used by large and developed economies.

8.6 Review Questions

1. Define tariff. Discuss the types of tariff.
2. Explain Non-tariff barriers and its types.
3. Distinguish between quotas and non-teriff barriers.

Answers: Self-Assessment

1. (*i*) (*a*)   (*ii*) (*d*)   (*iii*) (*b*)   (*iv*) (*c*)   (*v*) (*b*)
   (*vi*) (*a*)   (*vii*) (*b*)

8.7 Further Readings

Books
1. Evans, G., Newnham, J., Dictionary of International Relations; Penguin Books, 1998.
2. Filanlyason, J., Zakher M., The GATT and the regulation of Trade Barriers: Regime Dynamic and Functions; International Organization, Vol. 35, No. 4, 1981.
3. Frieden, J., Lake, D., International political economy: perspectives on global power and wealth, London: Routledge, 1995.
4. Mansfield, E., Busch, M., The political economy of Non-tariff barriers: a cross national analysis; International Organization, Vol. 49, No. 4, 1995.
5. Oatley,T., International political economy: interests and institutions in the global economy; Harlow: Longman, 2007.

Unit 9: Economic Effects of Tariff and Quotas on National Income, Output and Employment

---

CONTENTS

Objectives

Introduction

9.1 Impact to the Economy of a Country with the Tariff imposed on it

9.2 Empirical Evidence on the Effect of Tariffs

9.3 Summary

9.4 Key-Words

9.5 Review Questions

9.6 Further Readings

Objectives

After reading this Unit students will be able to:

• Analyse the Economic Effects of Tariff and Quotas

• Discuss Impact to the Economy of A country with the Tariff Imposed on It.

Introduction

A tariff is simply a tax or duty placed on an imported good by a domestic government. Tariffs are usually levied as a percentage of the declared value of the good, similar to a sales tax. Unlike a sales tax, tariff rates are often different for every good and tariffs do not apply to domestically produced goods.

The upcoming book Advanced International Trade: Theory and Evidence by Robert Feenstra gives three situations in which governments often impose tariffs:

• To protect fledgling domestic industries from foreign competition.

• To protect aging and inefficient domestic industries from foreign competition.

• To protect domestic producers from dumping by foreign companies or governments. Dumping occurs when a foreign company charges a price in the domestic market which is "too low". In most instances "too low" is generally understood to be a price which is lower in a foreign market than the price in the domestic market. In other instances "too low" means a price which is below cost, so the producer is losing money.

The cost of tariffs to the economy is not trivial. The World Bank estimates that if all barriers to trade such as tariffs were eliminated, the global economy would expand by 830 billion dollars by 2015. The economic effect of tariffs can be broken down into two components:

• The impact to the country which has a tariff imposed on it.

• The impact to the country imposing the tariff.

In almost all instances the tariff causes a net loss to the economies of both the country imposing the tariff and the country the tariff is imposed on.

9.1 Impact to the Economy of a Country with the Tariff Imposed on It

It is easy to see why a foreign tariff hurts the economy of a country. A foreign tariff raises the costs of domestic producers which causes them to sell less in those foreign markets. In the case of the softwood

lumber dispute, it is estimated that recent American tariffs have cost Canadian lumber producers 1.5 billion Canadian dollars. Producers cut production due to this reduction in demand which causes jobs to be lost. These job losses impact other industries as the demand for consumer products decreases because of the reduced employment level. Foreign tariffs, along with other forms of market restrictions, cause a decline in the economic health of a nation. The next section explains why tariffs also hurt the economy of the country which imposes them.

Except in all but the rarest of instances, tariffs hurt the country that imposes them, as their costs outweigh their benefits. Tariffs are a boon to domestic producers who now face reduced competition in their home market. The reduced competition causes prices to rise. The sales of domestic producers should also rise, all else being equal. The increased production and price causes domestic producers to hire more workers which causes consumer spending to rise. The tariffs also increase government revenues that can be used to the benefit of the economy.

There are costs to tariffs, however, now the price of the good with the tariff has increased, the consumer is forced to either buy less of this good or less of some other good. The price increase can be thought of as a reduction in consumer income. Since consumers are purchasing less, domestic producers in other industries are selling less, causing a decline in the economy.

Generally the benefit caused by the increased domestic production in the tariff protected industry plus the increased government revenues does not offset the losses the increased prices cause consumers and the costs of imposing and collecting the tariff. We haven't even considered the possibility that other countries might put tariffs on our goods in retaliation, which we know would be costly to us. Even if they do not, the tariff is still costly to the economy. In my article The Effect of Taxes on Economic Growth we saw that increased taxes cause consumers to alter their behavior which in turn causes the economy to be less efficient. Adam Smith's The Wealth of Nations showed how international trade increases the wealth of an economy. Any mechanism designed to slow international trade will have the effect of reducing economic growth. For these reasons economic theory teaches us that tariffs will be harmful to the country imposing them. That's how it should work in theory. How does it work in practice?

9.2 Empirical Evidence on the Effect of Tariffs

Study after study has shown that tariffs cause reduced economic growth to the country imposing them. A few of examples:

1. The essay on Free Trade at The Concise Encyclopedia of Economics looks at the issue of international trade policy. In the essay, Alan Blinder states that "one study estimated that in 1984 U.S. consumers paid $42,000 annually for each textile job that was preserved by import quotas, a sum that greatly exceeded the average earnings of a textile worker. That same study estimated that restricting foreign imports cost $105,000 annually for each automobile worker's job that was saved, $420,000 for each job in TV manufacturing, and $750,000 for every job saved in the steel industry."

2. In the year 2000 President Bush raised tariffs on imported steel goods between 8 and 30 percent. The Mackinac Center for Public Policy cites a study which indicates that the tariff will reduce U.S. national income by between 0.5 to 1.4 billion dollars. The study estimates that less than 10,000 jobs in the steel industry will be saved by the measure at a cost of over $400,000 per job saved. For every job saved by this measure, 8 will be lost.

The cost of protecting these jobs is not unique to the steel industry or to the United States. The National Center For Policy Analysis estimates that in 1994 tariffs cost the U.S. economy 32.3 billion dollars or $170,000 for every job saved. Tariffs in Europe cost European consumers $70,000 per job saved while Japanese consumers lost $600,000 per job saved through Japanese tariffs.

These studies, like many others, indicate that tariffs do more harm than good. If these tariffs are so bad for the economy, why do governments keep enacting them? We'll discuss that question in the next section.

Study after study has shown that tariffs, whether they be one tariff or hundreds, are bad for the economy. If tariffs do not help the economy, why would a politician enact one? After all politicans

are reelected at a greater rate when the economy is doing well, so you would think it would be in their self interest to prevent tariffs.

Recall that tariffs are not harmful for everyone, and they have a distributive effect. Some people and industries gain when the tariff is enacted and others lose. The way gains and losses are distributed is absolutely crucial in understanding why tariffs along with many other policies are enacted. To understand the logic behind the policies we need to understand The Logic of Collective Action. My article titled The Logic of Collective Action discusses the ideas of a book by the same name, written by Mancur Olson in 1965. Olson explains why economic policies are often to the benefit of smaller groups at the expense of larger ones. Take the example of tariffs placed on imported Canadian softwood lumber. We'll suppose the measure saves 5,000 jobs, at the cost of $200,000 per job, or a cost of 1 billion dollars to the economy. This cost is distributed through the economy and represents just a few dollars to every person living in America. It is obvious to see that it's not worth the time and effort for any American to educate himself about the issue, solicit donations for the cause and lobby congress to gain a few dollars. However, the benefit to the American softwood lumber industry is quite large. The ten-thousand lumber workers will lobby congress to protect their jobs along with the lumber companies that will gain hundreds of thousands of dollars by having the measure enacted. Since the people who gain from the measure have an incentive to lobby for the measure, while the people who lose have no incentive to spend the time and money to lobby against the issue, the tariff will be passed although it may, in total, have negative consequences for the economy.

You can meet the workers whose jobs will be lost if tariffs are not enacted by the government. Since the costs of the policies are distributed far and wide, you cannot put a face on the cost of a poor economic policy. Although 8 workers might lose their job for every job saved by a softwood lumber tariff, you will never meet one of these workers, because it is impossible to pinpoint exactly which workers would have been able to keep their jobs if the tariff was not enacted. If a worker loses his job because the performance of the economy is poor, you cannot say if a reduction in lumber tariffs would have saved his job. The nightly news would never show a picture of a California farm worker and state that he lost his job because of tariffs designed to help the lumber industry in Maine. The link between the two is impossible to see. The link between lumber workers and lumber tariffs is much more visible and thus will garner much more attention.

The gains from a tariff are clearly visible but the costs are hidden, it will often appear that tariffs do not have a cost. By understanding this we can understand why so many government policies are enacted which harm the economy. If you'd like to ask a question about tariffs, taxation, international trade or any other topic or comment on this story.

Did u know? The gains from tariff policies are a lot more visible than the losses. You can see the saw mills which would be closed down if the industry is not protected by tariffs.

Self-Assessment

1. **Choose the correct options:**
   (*i*) Over the past several years, the volume of world trade has:
       (*a*) decreased because of increased government regulation of trade through tariffs and quotas
       (*b*) decreased because of the decline in incomes brought on by global recession
       (*c*) increased because of increased government regulation of trade through tariffs and quotas
       (*d*) increased because of improvements in communications and transportation
   (*ii*) A tariff:
       (*a*) raises the price of imported goods, increasing the demand for domestic substitutes
       (*b*) lowers the cost of producing domestic goods
       (*c*) offsets the effect of a quota
       (*d*) raises the price of domestic goods, lowering the demand for them

(*iii*) U.S. exports:

    (*a*) average 4-6 percent of GDP, as do imports

    (*b*) average 18-20 percent of GDP, while imports average 10-12 percent

    (*c*) average 10-12 percent of GDP, while imports average 15-17 percent

    (*d*) average 1-2 percent of GDP, while imports average 4-6 percent

(*iv*) Two nations, Gamma and Delta, both produce shoes and dresses. Gamma has a comparative advantage in the production of shoes if:

    (*a*) it can produce shoes with fewer resources than Delta

    (*b*) its domestic opportunity cost of shoes in terms of dresses is lower than Delta's

    (*c*) its supply of shoes is greater than Delta's

    (*c*) it is wealthier than Delta

(*v*) A decrease in the U.S. demand for Mexican goods will:

    (*a*) increase the demand for the peso and increase its dollar price

    (*b*) increase the supply of the peso and decrease its dollar price

    (*c*) decrease the supply of the peso and increase its dollar price

    (*d*) decrease the demand for the peso and decrease its dollar price

(*vi*) One major outcome of the North American Free Trade Agreement is:

    (*a*) massive investment by Asian companies in Mexico to exploit reduced tariffs

    (*b*) increased unemployment in Mexico

    (*c*) higher average living standards in Canada, Mexico, and the U.S.

    (*d*) reduced exports from the U.S. to Mexico and Canada

9.3 Summary

- A tariff is simply a tax or duty placed on an imported good by a domestic government. Tariffs are usually levied as a percentage of the declared value of the good, similar to a sales tax. Unlike a sales tax, tariff rates are often different for every good and tariffs do not apply to domestically produced goods.

- The cost of tariffs to the economy is not trivial. The World Bank estimates that if all barriers to trade such as tariffs were eliminated, the global economy would expand by 830 billion dollars by 2015. The economic effect of tariffs can be broken down into two components.

- It is easy to see why a foreign tariff hurts the economy of a country. A foreign tariff raises the costs of domestic producers which causes them to sell less in those foreign markets.

- There are costs to tariffs, however, now the price of the good with the tariff has increased, the consumer is forced to either buy less of this good or less of some other good. The price increase can be thought of as a reduction in consumer income. Since consumers are purchasing less, domestic producers in other industries are selling less, causing a decline in the economy.

- Generally the benefit caused by the increased domestic production in the tariff protected industry plus the increased government revenues does not offset the losses the increased prices cause consumers and the costs of imposing and collecting the tariff. We haven't even considered the possibility that other countries might put tariffs on our goods in retaliation, which we know would be costly to us. Even if they do not, the tariff is still costly to the economy.

9.4 Key-Words

1. Tariff policies   : A tariff policy is a strategy of taxing imported or exported goods and services from one country to another. These taxes often seek to protect domestic industries or punish countries for policies related or unrelated to the economy. Considered by friendlier countries to be a barrier to

production and positive interaction, some countries have created exceptions to their tariff policies known as free trade agreements or free trade zones.

2. Empirical Evidence : Empirical evidence (also empirical data, sense experience, empirical knowledge, or the a posteriori) is a source of knowledge acquired by means of observation or experimentation. Empirical evidence is information that justifies a belief in the truth or falsity of an empirical claim. In the empiricist view, one can only claim to have knowledge when one has a true belief based on empirical evidence.

9.5 Review Questions

1. Discuss the Impact to the Economy of a Country with the Tariff Imposed on It.
2. What is the Empirical Evidence on the Effect of Tariffs? Discuss

Answers: Self-Assessment

1. (*i*) (*d*)   (*ii*) (*a*)   (*iii*) (*c*)
    (*iv*) (*b*)   (*v*) (*d*)   (*v*) (*c*)

9.6 Further Readings

1. Evans, G., Newnham, J., Dictionary of International Relations; Penguin Books, 1998.
2. Filanlyason, J., Zakher M., The GATT and the regulation of Trade Barriers: Regime Dynamic and Functions; International Organization, Vol. 35, No. 4, 1981.
3. Frieden, J., Lake, D., International political economy: perspectives on global power and wealth, London: Routledge, 1995.
4. Mansfield, E., Busch, M., The political economy of Non-tariff barriers: a cross national analysis; International Organization, Vol. 49, No. 4, 1995.
5. Oatley,T., International political economy: interests and institutions in the global economy; Harlow: Longman, 2007.

Unit 10 : Political Economy of Non-tariff Barriers : and Their Applications

---

CONTENTS

Objectives

Introduction

Objectives

After reading this Unit students will be able to:

• Analyse   the Political Economy of Non-tariff Barriers.

• Discuss the Societal Approaches to Trade Policy.

Introduction

Much research on the determinants of trade policy has focused on the efficacy of societal and statist approaches. Societal theories typically attribute patterns of protection to variations in demands made by pressure groups, whereas statist theories emphasize the effects of the "national interest" and domestic institutions in determining the level of protection. While both approaches have gained considerable currency, debates concerning their relative merits have been heated and long-standing. Yet very little quantitative evidence has been brought to bear on this topic.

In this article, we provide some of the first results of this sort. Our findings indicate that although societal and statist approaches often are considered mutually exclusive, it is more fruitful to view them as complementary. Moreover, the interaction between factors that give rise to demands for protection and those that regulate the provision of protection by policymakers has not been treated adequately in the literature on foreign economic policy. This gap in the literature is fundamentally important, since our results indicate that the interaction between these factors is a central determinant of trade policy. Thus, analyses of commercial policy that fail to consider both societal and statist variables and the interaction between them are likely to be inadequate.

Our analysis centers on explaining cross-national patterns of nontariff barriers (NTBs). Scholars have conducted little cross-national research on trade policy and virtually none with a focus on NTBs. Instead, single-country studies of tariffs comprise much of the existing literature on the political economy of commercial policy. Yet the usefulness of societal and statist theories of foreign economic policy hinges on the ability of these theories to explain variations in protection across states, and NTBs have become increasingly pervasive among the advanced industrial countries. Because the General Agreement on Tariffs and Trade (GATT) and the World Trade Organization (WTO) limit the ability of contracting parties to impose tariffs, policymakers who view protection as an attractive means by which to meet the demands of pressure groups or advance state interests are likely to rely primarily on NTBs. Many observers have suggested that this is occurring with increasing regularity and that the recent proliferation of NTBs has done much to offset the gains in liberalization made

during successive rounds of the GATT. A fuller understanding is therefore needed of the factors that account for variations in NTBs across states.

10.1 Societal Approaches to Trade Policy

Societal (or pluralist) approaches to the study of foreign economic policy focus primarily on the effects of demands for protection by pressure groups. Societal explanations consider trade policy to be the product of competition among pressure groups and other nonstate actors that are affected by commerce. The impact of these groups on policy depends largely on their ability to organize for the purpose of articulating their demands and on the amount of electoral influence they possess. Societal approaches attribute little importance to policymakers and political institutions for the purposes of explaining trade policy. As G. John Ikenberry, David Lake, and Michael Mastanduno point out, societal theories view the state as "essentially passive; it acts as a disinterested referee for competing groups, and supplies policies to satisfy the demands of successful domestic players."

Societal approaches to the study of trade policy characterize much of the literature on endogenous protection. Empirical studies of this sort infer the demands for protection based on macroeconomic and/or sectoral fluctuations. Most analyses of endogenous protection conducted by political scientists have been cast at the sectoral level. A large and growing body of literature, however, centers on the macroeconomic determinants of protection. Much of this research supports the view advanced by certain societal theories that macroeconomic fluctuations strongly influence pressures for protection. Therefore we focus our societal analysis of NTBs on macroeconomic factors.

Chief among the macroeconomic variables that these studies emphasize are unemployment and the real exchange rate. It is widely accepted by analysts of trade policy that high levels of unemployment contribute to demands for protection. Widespread unemployment increases the costs to workers of adjusting to rising import levels. Workers who are displaced by imports will find it progressively more difficult to obtain alternative employment, and when they do, downward pressure will be placed on their wages. Together these factors promote pressures to restrict the flow of imports.

In addition to unemployment, variations in the exchange rate are expected to give rise to protectionist pressures. In fact, Rudiger Dornbusch and Jeffrey Frankel argue that "hypotheses concerning the exchange rate may be the most important macroeconomic theories of protection."

Did u know? "Conventional wisdom suggests that high levels of unemployment are the single most important source of protectionist pressures."

Central to the effects of the exchange rate on demands for protection is the influence of the price of a state's currency on the competitiveness of its exports and its import-competing products. An appreciated currency, by increasing the price of domestically produced goods, threatens to undermine both exports and import-competing sectors of the economy. As C. Fred Bergsten and John Williamson point out in a related context, these developments are likely to contribute to "pressure that is generated for protectionist measures. Export and import-competing firms and workers will tend to seek help from their governments to offset these distortions, which undermine their ability to compete, with some degree of legitimacy since the distortions are accepted—in some cases, even fostered—by those governments. Coalitions in support of trade restrictions will be much easier to form, and much broader in their political clout, because no longer will only the most vulnerable firms and workers be seeking help—and no longer will the countervailing pressures from successful exporters be as effective."

Public officials in liberal democracies are expected to meet demands for protection that arise due to high levels of unemployment and an appreciated currency because these variables influence the voting behavior of constituents. There is evidence that voters cast ballots on the basis of their personal economic circumstances, especially if they are recently unemployed. However, substantial evidence also indicates that voters cast ballots on the basis of macroeconomic conditions, regardless of whether they are directly affected by these conditions. In fact, some studies have concluded that macroeconomic

factors are more salient determinants of voting behavior than are personal economic circumstances. Other survey research further suggests that public support for protection increases during downturns in the economy and when domestic industries are under severe pressure from foreign competition. As a result, public officials seeking to enhance their electoral fortunes have incentives to impose protection during periods of high unemployment and currency appreciation because such measures are likely to be popular and may blunt the short-term effects of macroeconomic pressures. These analyses therefore lead us to expect a direct relationship to exist between both the level of unemployment and the real exchange rate, on the one hand, and the incidence of NTBs, on the other hand.

10.2 Statist Approaches to Trade Policy

While societal approaches have been especially influential in the field of political economy, they also have been criticized on a number of grounds. Especially important is the charge leveled by statists and others that societal approaches systematically underestimate the effects of two factors that regulate the provision of protection : state interests with respect to trade policy and domestic institutions. Analyses that emphasize state interests generally focus on the roles of politicians and policymakers in the formation of trade policy, holding constant societal pressures. Further, as Ikenberry, Lake, and Mastanduno note, these analyses presume that the preferences of public officials "are partially, if not wholly, distinct from the parochial concerns of either societal groups or particular government institutions, and are tied to conceptions of the 'national interest' or the maximization of some social welfare function."

Many statists conclude that the ability of policymakers to advance the national interest depends in large measure on the extent to which domestic political institutions render them susceptible to demands by pressure groups and other nonstate actors. Policymakers who are poorly insulated from, and lack autonomy with respect to, pressure groups will face difficulty advancing the national interest unless (as discussed further below) it converges with the preferences of societal groups. Thus, one hypothesis we will test is that institutional factors that foster the insulation and autonomy of public officials bolster the ability of states to pursue trade policy consistent with the national interest.

Relative size

Clearly, the national interest with respect to trade is likely to vary across states; and it is not possible to assess adequately the influence of institutional factors on trade policy from a statist perspective unless each state's interest can be specified. On this score, many analysts have argued that a state's economic size governs its national interest with respect to trade policy.

There is ample reason to expect that larger states will display a more pronounced interest in protection than their smaller counterparts. First, international trade theory suggests that this should be the case. By virtue of their size, large states are likely to be vested with disproportionate market power. They can exploit their monopoly power through the use of tariffs, as well as quotas and other NTBs that duplicate a tariffs effect. If the imposition of an optimal quota elicits retaliation, the welfare of both parties will suffer. This, however, only limits the incentives for a large state to impose NTBs against a state of similar size, since only states with some monopoly power have an incentive to retaliate in response to the imposition of protection. Large states retain an incentive to target small states, since the latter have no incentive to retaliate. In contrast, small states are unlikely to possess the market power necessary to benefit from optimal protection and face the prospect of retaliation by trade partners (thereby reducing foreign commerce on which they tend to be highly dependent) if they impose NTBs. Hence, on average, we expect larger states to display a greater preference for NTBs than their smaller counterparts.

Second, state size is likely to be directly related to patterns of protection due to the time period analyzed in this article. As discussed further below, the empirical analysis conducted in this study is based on the mid-1980s. In the opinion of many scholars, this was a period characterized by a moderately skewed distribution of power among a few relatively large nonhegemonic states. A number of studies have concluded that systems of this sort—as well as ones in which hegemony is declining—

provide incentives for the dominant states to behave in a commercially predatory manner. Based on these considerations, we expect that economic size will be directly related to the incidence of NTBs.

We measure a state's relative economic size in two ways : the ratio of its imports to total global imports and the ratio of its gross domestic product (GDP) to total global GDP. The first variable has been used repeatedly as a measure of economic size. The second is also important because states with relatively large GDPs are likely to possess greater market power and to be better able to forgo commerce than are states with relatively small GDPs.

Although it is clear that these two measures of relative size should be highly correlated, analyzing both allows us to determine whether our empirical results are sensitive to the measure that is used. Moreover, analyzing both measures of size is important because the ratio of national imports to global imports is closely related to the measure of trade dependence used by Ronald Rogowski; and his analysis implies that any observed effect on NTBs of those domestic institutions on which we focus here (and that are discussed below) might be due to the effects of trade dependence on both domestic institutions and NTBs. Further, including national GDP as a percentage of global GDP is important because Wendy Takacs links the level of national product to macroeconomic cycles that give rise to demands for protection. Contrary to the hypothesis discussed above, however, she finds that national product is inversely related to escape clause investigations and (to a lesser degree) to positive findings by the U.S. International Trade Commission in such cases. Including both measures of size allows us to examine each of these issues.

Domestic institutions

From a statist viewpoint, NTBs should be most prevalent in large states characterized by high degrees of institutional insulation and autonomy, since these conditions provide policymakers with an economic incentive to impose NTBs and vest them with the capacity to advance those interests.

Our analysis of institutions draws heavily on an important study by Rogowski. He argues that "insulation from regional and sectoral pressure in a democracy ... is most easily achieved with large electoral districts. ... [This argument is] easily defended, in part because institutional theorists have almost universally accepted it..., but more because it is almost self-evident. When automakers or dairy farmers entirely dominate twenty small constituencies and are a powerful minority in fifty more, their voice will be heard in a nation's councils. When they constitute but one or two percent of an enormous district's electorate, representatives may defy them more freely."

> *Notes* Rogowski maintains that the autonomy of public officials in democratic states is bolstered by both large constituencies and the existence of a list-system PR regime.

Rogowski therefore relies on the (natural logarithm) number of parliamentary constituencies in the most powerful legislative body (or, in those cases where the most powerful body is not obvious, the chamber with the most members) of each democratic state as a measure of institutional insulation and autonomy. It is expected that insulation and autonomy will be inverse functions of the number of constituencies in any democracy. All other things being equal, a larger number of constituencies reduces the size of the average constituency in each state. The smaller is this average size, the more homogeneous is each district, the fewer is the number of special interests that are likely to exist per constituency, and the greater will be the political influence of each pressure group in that district.

As a result, small electoral districts encourage patronage and pork-barrel politics. Since legislators representing small districts are likely to be beholden to a few influential pressure groups, they are likely to attempt to provide those groups with benefits, including trade policies that reflect their preferences. Yet in polities composed of many small constituencies, no single legislator has the capacity to provide these benefits. As Barry Weingast and William Marshall argue, "This, in combination with the diversity of interests they represent, generates a gain from exchange and cooperation among legislators." The logrolling to which this situation gives rise is likely to yield trade policy that covers

more types of goods and services than would be the case in a country characterized by large electoral districts and less influential interest groups.

In addition to the number of parliamentary constituencies, another important institutional feature of democracies concerns whether a list-system proportional representation (PR) or a winner-take-all system exists. Rogowski notes, "Pressure groups are restrained where campaign resources or the legal control of nominations are centralized in the hands of party leaders. Of course, such control is achieved quite effectively in rigid list-system PR."

We begin the following analysis by focusing on the effects of the number of parliamentary constituencies on NTBs. We then analyze the effects of both the number of constituencies and whether a state's electoral system is winner-take-all or PR. We take this tack to determine whether our results are robust with respect to the inclusion of PR in our model. Doing so is useful because some studies have found that differences exist in the autonomy of policymakers among various PR systems and that, in certain cases, public officials in PR systems may be vested with less autonomy than their counterparts in winner-take-all systems.

10.3 Effects of the Interaction between Societal and Statist Factors on Trade Policy

Although the relative merits of societal and statist approaches to the study of foreign economic policy have been hotly debated, quantitative comparisons of them have been rare. Rarer still have been empirical analyses of the impact of the interaction between societal and statist factors on trade policy. The few previous studies of this sort have focused almost exclusively on U.S. commercial policy. Additionally, much of their effort has been directed at explaining tariffs, rather than NTBs. However, its unusual size and institutional structure may limit how broadly conclusions based on studies of U.S. trade policy can be applied. Further, it is widely recognized that successive rounds of the GATT limited the ability of member states to respond to demands for protection or to advance state interests by imposing tariffs. Since the states we examine are all contracting parties to the GATT, our focus on explaining NTBs (rather than tariffs) is especially important. Indeed, the need to analyze the effects of factors that govern demands for and the provision of protection—as well as the interactions among them—on cross-national patterns of NTBs is a glaring gap in the literature, and one that we seek to redress.

We focus on two related issues concerning the interaction between societal and statist factors. As noted above, some statists argue that the policies of states in which policymakers are poorly insulated from societal pressures tend to reflect the interests of societal groups rather than the national interest. There is also reason to expect increases in societal demands for protection during cyclical downturns in the economy and when macroeconomic conditions undermine the competitiveness of a state's goods. Thus, one hypothesis we will examine is that the incidence of NTBs tends to be greatest in states characterized by (1) high levels of unemployment and appreciated currencies and (2) domestic institutions that undermine the insulation and autonomy of public officials with respect to pressure groups.

Notes What you mean by Statist Factor?

A second hypothesis we will test is that the incidence of NTBs is greatest in cases where both state and societal actors display a preference for protection. Although much of the literature bearing on the interaction between statist and societal factors has focused on the effects of divergent preferences between state and societal actors, the possibility that their preferences might converge has been raised in a number of studies, although it has not been analyzed in sufficient detail. If such a convergence is

an important determinant of NTBs, then their incidence should be greatest in large states characterized by (1) high levels of unemployment and appreciated currencies and (2) political institutions that bolster the insulation and autonomy of public officials with respect to pressure groups. As noted above, deteriorating macroeconomic conditions elicit demands for protection, and public officials who fail to respond to these demands may suffer accordingly in subsequent elections. Further, in contrast to small states, large states often have an incentive to impose protection; and public officials that are well-insulated and vested with considerable autonomy will be in a position to act on those incentives, and would be expected to do so.

A high degree of institutional insulation and autonomy is essential in this regard. Although we expect high levels of unemployment and appreciated currencies to yield widespread demands for protection, some societal groups are likely to retain an interest in lower trade barriers. These groups include multinational corporations, industries that depend on or are highly sensitive to the price of imports, and industries that depend on exports and fear either that increases in protection by their government will elicit retaliation by foreign governments or that protection will reduce foreign exports and hence the ability of foreign consumers to purchase their imports. Moreover, in a study of U.S. trade policy, I. M. Destler and John Odell found that the political pressure exerted by these antiprotectionist forces increased during those periods when macroeconomic downturns led to broad-based societal pressures for protection. Their influence, like that of other societal groups, depends on the structure of domestic institutions. Thus, large states characterized by high levels of unemployment and appreciated currencies should experience a higher incidence of NTBs when institutions insulate policymakers from those groups that prefer lower trade barriers than when porous institutions enhance the influence of these groups on trade policy.

10.4 The Relationship between Tariffs and NTBs

In addition to the hypotheses described above, we also examine the effects of preexisting tariff levels on NTBs. Doing so is important because preexisting tariff levels may influence both the strength of societal demands for NTBs and the willingness of public officials to meet these demands. Groups already well protected by tariffs may bring less pressure for new NTBs and face more governmental resistance to their demands than less well protected groups. This suggests that tariffs and NTBs are substitutes, which is consistent with the view expressed by some economists that NTBs are often used to protect industries that have lost tariff protection due to successive rounds of the GATT. Jagdish Bhagwati refers to this dynamic as the "law of constant protection." As he points out, "The evidence of increased nontariff barriers and administered protection just as tariffs had been reduced to new lows suggests the intriguing possibility that there may be a Law of Constant Protection : If you reduce one type of protection, another variety simply pops up elsewhere. (You then have a Displacement Effect, not evidence of any increase in protectionist pressure.)"

In contrast to this view, another prominent position holds that tariffs and NTBs are complements. Those who advance this argument maintain that NTBs are often used to protect those industries that are also the beneficiaries of high tariffs, while states avoid using NTBs to shield industries that receive little tariff protection. Edward John Ray, for example, mentions that U.S. NTBs may be concentrated in industries least affected by the Kennedy Round of the GATT. In contrast to the law of constant protection, a direct relationship between tariffs and NTBs might suggest that NTBs are used to counter new foreign challenges to important sectors that are already the beneficiaries of tariff protection. Indeed, the results of a number of single-country analyses seem to support this position. Cross-national studies, however, have produced more ambiguous evidence on this score.

A related reason to include tariffs in our model is that they might account for any observed relationship between societal and statist variables, on the one hand, and the incidence of NTBs, on the other hand. Various studies have found that the unemployment rate, the exchange rate, economic size, and institutional factors are related to patterns of tariffs; and the research discussed in this section links tariffs to patterns of NTBs. It is therefore important to determine whether tariffs influence the effects of macroeconomic and institutional factors on NTBs.

A model of non-tariff barriers to trade

Our initial model, then, is :

$$NTB_{t+1} = A + B_1 SIZE_t + B_2 (\log CONST)_t + B_3 (SIZE \cdot \log CONST)_t + (1)$$

$$B_4 UNEM_t + B_5 (UNEM \cdot \log CONST)_t + B_6 (UNEM \cdot SIZE \cdot \log CONST)_t$$

$$+B_7 REER_t + B_8 (REER \cdot \log CONST)_t + B_9$$

$$(REER \cdot SIZE \cdot \log CONST)_t + B_{10} TARIFF_t + e_t.$$

The dependent variable, NTB_{t+1}, is the proportion of imports subject to NTBs in each state in year t + 1 based on the United Nations Conference on Trade and Development's (UNCTAD) "inventory list" of NTBs. As Sam Laird and Alexander Yeats observe, this list includes "Variable import levies and product specific charges (excluding tariff quotas); Quotas; Prohibitions (including seasonal prohibitions) : non-automatic import authorisations including restrictive import licensing requirements; quantitative 'voluntary' export restraints; and trade restraints under the Multifibre Arrangement." It is, of course, difficult to gauge the extent of NTB protection within or across states. Unlike tariffs, NTBs have no natural measure of intensity, and calculations of advalorem equivalents of NTBs have proven to be unreliable. Some analysts have treated NTBs as a dichotomous variable based on whether or not a given sector is protected by them. Others have estimated the level of NTBs based on the difference between the predicted and observed openness of an economy.

In contrast to these approaches, we examine the incidence of NTBs. This measure is chosen because the UNCTAD trade coverage ratios are viewed by many experts as the most reliable estimates of NTBs across states and because it is the most appropriate variable with which to test our theory. For example, polities characterized by many (and therefore small) parliamentary constituencies may be especially prone to pork-barrel politics. Under these conditions, logrolling is likely to be pervasive and the preferences of many different interest groups are therefore likely to be reflected in trade policy. Since the extent and variety of interest-group demands reflected in trade policy bear directly on the incidence of protection, we focus on explaining the incidence of NTBs. The coverage ratios that we analyze measure the proportion of a state's imports that are subject to NTBs.

Turning to the independent variables, $SIZE_t$ is the economic size of each state in year t. It was pointed out earlier that both the ratio of national imports to total global imports and the ratio of national GDP to global GDP are used to measure relative state size. Thus, we estimate the parameters in equation (1) using both variables. In addition, $\log CONST_t$ is the natural logarithm of the number of parliamentary constituencies in each state in year t based on Rogowski's data; $UNEM_t$ is the unemployment rate in each state in year t; $REER_t$ is an index of the real exchange rate in each state in year t; $TARIFF_t$ is the average national post-Tokyo Round offer rate for each state; and e_t is an error term. The remaining variables are included in order to determine whether, as we hypothesized above, the interaction between factors that regulate the provision of protection (SIZE$\cdot$ log CONST*)*, and the interaction between factors that govern demands for protection and those that regulate its supply, are important determinants of cross-national patterns of NTBs.

Data limitations led us to focus on explaining NTBs in 1983 and 1986. UNCTAD provides data on NTBs for fourteen advanced industrial states in these years. The fourteen states are : Belgium-Luxembourg, Denmark, Finland, France, Greece, Ireland, Italy, Japan, the Netherlands, Norway, Switzerland, the United Kingdom, the United States, and West Germany. Although UNCTAD also provides NTB data for New Zealand, our results indicated that this country was a statistical outlier. We therefore excluded New Zealand and focus on the aforementioned fourteen states in the following analysis. Because NTBs are measured in 1983 and 1986 (years t + 1), the independent variables in equation (1) (except for TARIFF*)* are measured in 1982 and 1985 (years t). The observations for 1983 and 1986 are initially pooled; however, the extent to which the incidence of NTBs varied between 1983 and 1986 is also examined below.

It should be noted at the outset that all of the states analyzed in this study are advanced industrial countries. This precludes, for example, an assessment of whether our findings vary depending on a state's level of economic development. It is also clear that caution must be exercised when offering generalizations based on an analysis of such a limited time period. But since the tendency for advanced industrial countries to rely on NTBs became increasingly pervasive during the 1980s and virtually no quantitative cross-national research has been conducted on the issues addressed here, our results should provide a useful first cut at the hypotheses presented above.

Regression diagnostics

Before proceeding, a number of issues regarding the regression results presented need to be addressed. First, in a cross-sectional analysis such as that conducted here, one concern is that the errors in the regression (e_t) will not have a common variance. Under these circumstances, the OLS estimates will be heteroscedastic and therefore inefficient. White tests yielded no evidence of heteroscedasticity in the present case.

Second, as noted above, it is important to ensure that our decision to pool data across 1983 and 1986 was appropriate. It is obvious that if the effects of the independent variables in equations (1) and (2) on NTBs vary over time, this procedure would be inappropriate. However, analysis-of-variance tests yielded no evidence that the model is unstable between 1983 and 1986, and Chow tests also yielded no evidence that any of the individual regression coefficients is unstable across time.

Third, the use of OLS might be inappropriate because the value of the dependent variable (NTB) is a proportion (and therefore is bounded by zero and one). Under these conditions, OLS estimates may be inefficient and predict proportions of NTB that exceed one or are less than zero. Since it is well-known that a Tobit model can be used to deal with these problems, we estimated the parameters in equations (1) and (2) using that model. The results were virtually identical to those. Apparently, the fact that the dependent variable is truncated poses no problem here.

Fourth, we should examine the extent to which our results are sensitive to influential observations. To this end, we estimated the parameters in equations (1) and (2) after deleting observations one at a time. Our findings indicated that the signs and statistical significance of the regression coefficients are quite robust with respect to the deletion of individual observations. Regardless of which measure of economic size was used, there was no case in which the sign of a regression coefficient changed, and in fewer than 5 percent of the cases did a regression coefficient fail to remain significant once any observation was removed from the sample. Indeed, this number of changes would be expected by chance alone.

Fifth, it is important to ensure that our results are not vitiated by multicollinearity. The results of bivariate correlations and auxiliary regressions indicated that all of those few cases in which multicollinearity might be a problem involved PR and TARIFF. To determine whether the effects of the remaining variables in the model were sensitive to the inclusion of PR and TARIFF, we deleted these variables (as well as PR · SIZE) from the model individually and in combination. The signs, sizes, and levels of statistical significance of the remaining variables in equations (1) and (2) were quite robust with respect to the inclusion or omission of these variables.

Finally, it is useful to consider the possibility that variables omitted from our model may influence the findings. Particularly important in this regard is whether the extent of government intervention in the economy influences the rate of unemployment and its propensity to impose NTBs. We therefore included in equations (1) and (2) the ratio of government expenditures to GDP in year t, which is a measure of government intervention. The results of this analysis indicate that the regression coefficient of this variable is negative and statistically significant in every case. This may reflect the tendency for states characterized by high levels of government spending to buffer and compensate societal groups disproportionately, thereby reducing demands for protection from groups that are adversely affected by imports. It is important to note, however, that the inclusion of this variable in our models has no effect on the sign or level of statistical significance of any remaining variable, including UNEM, UNEM· log CONST, and UNEM· SIZE· log CONST. Nor are the quantitative effects of the variables in equations (1) and (2)—and, hence, the conditions that maximize and minimize the incidence of NTBs—influenced in any substantial way by its inclusion.

Some illustrations of the statistical findings

Having tested our model, it is useful to illustrate how the societal and statist variables on which we focused affected trade policy in the countries considered here. While detailed case studies are beyond the scope of this article, anecdotal evidence suggests that these variables were salient influences on commercial policy during the 1980s.

Consider, for example, the role that the exchange rate played in United States trade policy. Between 1983 and 1986, the incidence of U.S. NTBs rose by over 25 percent. Much of this rise seems to be due to a significant appreciation in the dollar. While the values of the other independent variables in our model changed relatively little from 1982 to 1985 in the case of the United States, the value of REER increased dramatically. The societal view that this appreciation should precipitate an increase in demands for protection accords with a number of accounts of exchange-rate politics in the United States during this period. I. M. Destler and C. Randall Henning note that, by the early 1980s, many sectors of U.S. industry had concluded that the dollar's strength was degrading their competitiveness. By 1985, their opposition to the dollar's strength reached a peak. Imports were flooding into the United States at a rate unprecedented during the post-World War II era and Destler and Henning maintain that "most [industries] considered the prime source of their problems [to be] the sky-high dollar." They also argue that "[w]hen the strong dollar triggered a flood of imports, a rise in protectionist bills was a predictable result." U.S. industry and labor petitioned the Reagan administration and Congress to remedy the dollar appreciation. With no direct control over exchange-rate policy, Congress responded by "generat[ing] a veritable explosion of trade legislation initiatives" in 1985. It is also noteworthy that the period from 1982 to 1985 witnessed a rapid surge in the number of petitions for trade-policy relief by U.S. industry and a turn toward managed-trade policies by the United States. These developments both led directly to an increase in the incidence of NTBs and, in the opinion of J. David Richardson, were largely attributable to the dollar's appreciation.

The effects of unemployment on NTBs are illustrated by the case of West Germany during the 1980s. From 1983 to 1986 the incidence of West German NTBs rose by approximately 15 percent; and from 1982 to 1985, the level of West German unemployment rose by about 25 percent, while the remaining independent variables in our model experienced only very modest fluctuations. Kathleen Thelen points out that throughout the post–World War II era, "the [West German] government sought first and foremost to maintain a stable currency and hold inflation in check, even if it meant higher unemployment." By 1983, the West German economy had deteriorated to the point where unemployment had reached its highest level since the end of World War II. Of particular importance for present purposes was the structural nature of West German unemployment. In 1983, over a quarter of those West Germans without jobs had been unemployed for more than one year. The Bundesbank reported that "[t]he prospects for bringing unemployment down quickly to a more bearable level are admittedly slim; this will certainly not happen in the short term." Labor problems reached a peak in 1984 with the metal workers' strike, which was designed in part to reduce unemployment. It is interesting that the Organization for Economic Cooperation and Development reported in 1986 that West German NTBs were most pervasive in those sectors where tariffs had been reduced and that among these were sectors in which metal workers were employed in large numbers (such as steel). This suggests that the government responded to mounting unemployment by increasing the incidence of NTBs in 1986. Given the political strength of organized labor, the traditional unwillingness of the government to enact macroeconomic policies to counter unemployment at the risk of undermining monetary stability, and Germany's mounting unemployment problems, West Germany's course of action is not surprising.

Further, it is interesting to compare the effects of institutional variations between Japan and the United States on their respective propensities to impose NTBs. It is often argued that Japan is a "strong" state in which policymakers are extremely well-insulated and autonomous with respect to interest groups. The United States, on the other hand, is often portrayed as a "weak" state in which policymakers lack both insulation and autonomy. Yet both of these countries are characterized by a relatively large number of parliamentary constituencies based on our sample of states. This suggests that public officials in both countries are likely to be susceptible to societal pressures (although not necessarily to the same extent); and it jibes with the view expressed in a number of recent studies that

Japanese policymakers are far less autonomous and less insulated from interest groups than is implied by those who characterize it as a strong state. As Kent Calder argues, "Japan's system of medium-size electoral districts forces as many as five members of the largest political parties ... to run against one another in the same electoral district.... As a result,... legislators tend to be highly sensitive to constituency pressure, especially from relatively well-organized grassroots pressure groups such as agriculture and small business." A primary foreign policy interest of these groups is the prevention of the loss of domestic markets to imports, and this has led them to form alliances with politicians and bureaucrats that are likely to undermine the insulation and autonomy of these state actors.

Going a step further, it is useful to consider the results presented in light of this discussion of Japanese and U.S. institutions. In 1986, for example, Japan and the United States were the two largest states in our sample, both countries had appreciated currencies and relatively little unemployment, and neither state's electoral system was PR. From the standpoint of our model, the primary difference between them was that the United States had noticeably more constituencies than did Japan. As a result, it would be expected on the basis of this model that Japanese policymakers would be somewhat better insulated and more autonomous than their American counterparts, and that this institutional feature would better enable them to pursue the national interest. It, therefore, is not surprising that the incidence of NTBs was greater in Japan than in the United States during 1986. At the same time, however, both Japanese and U.S. NTBs were relatively high in 1986 based on the sample of countries considered here. This is consistent with the view described above that, while Japanese policymakers are vested with greater institutional capacity (and therefore are better able to advance the national interest as they see it) than their American counterparts, the institutional characteristics of Japan and the United States are more similar than is often recognized. Clearly, the cases presented in this section can be taken as no more than suggestive of the ways in which societal and statist factors influence trade policy. Yet these examples do illustrate why the variables emphasized in our model are so strongly related to cross-national patterns of NTBs.

Self-Assessment

1. **Choose the correct option**

   (*i*) Protecting domestic producers against import competition:

      (*a*) Helps those producers.

      (*b*) Helps domestic consumers of the product.

      (*c*) Probably helps the importing nation as a whole.

      (*d*) All of the above.

   (*ii*) A nontariff barrier operates by:

      (*a*) Limiting the quantity of imports.

      (*b*) Increasing the cost of getting imports to market.

      (*c*) Creating uncertainty about the conditions under which imports will be permitted.

      (*d*) All of the above.

   (*iii*) One of the reasons that protectionists and government officials may favor using a quota instead of a tariff is:

      (*a*) Quotas generate more revenue for the government than do tariffs.

      (*b*) A quota ensures that the quantity of imports is strictly limited.

      (*c*) Quotas create less market distortions than do tariffs.

      (*d*) Quotas give less power to politicians than do tariffs.

   (*iv*) A quota:

      (*a*) Causes domestic prices to fall.

      (*b*) Causes world prices to rise.

      (*c*) Restricts the quantity of a good that can be imported.

      (*d*) Is always more efficient than a tariff.

(*v*) In the case of a small country, a quota and a tariff are (almost) identical if:

    (*a*) The government allocates licenses for free to importers using a rule or process that involves (almost) no resource cost.

    (*b*) The government auctions off licenses to the highest bidder.

    (*c*) The government allocates licenses to importers through application and selection procedures that require the use of substantial resources.

    (*d*) The government allocates import licenses directly to the public using a free lottery system.

(*vi*) Which of the following is a means of allocating import licenses by assigning the licenses without competition, applications, or negotiation?

    (*a*) Fixed favoritism.

    (*b*) Resource-using application procedures.

    (*c*) Import-license auctions.

    (*d*) Domestic content requirements.

(*vii*) Which of the following requires that an import distributor buy a certain percentage of the product locally?

    (*a*) An import quota.

    (*b*) A mixing requirement.

    (*c*) A voluntary export restraint.

    (*d*) A domestic content requirement.

10.5 Summary

- Our results have a number of implications for studies of the political economy of trade policy. In recent years, one of the most persistent sources of debate among both economists and political scientists has centered on the relative merits of societal and statist explanations of foreign economic policy. Our findings lend support for the societal argument that macroeconomic fluctuations contribute to demands for protection, which are in turn central determinants of trade policy.

- Large states have a greater incentive to impose protection than their smaller counterparts, and our findings indicate that they do in fact impose NTBs more widely than small states. It is curious that, despite the clear importance of this factor, it has been considered so rarely in empirical research on trade policy. Our results indicate that this omission is likely to yield incomplete and potentially misleading conclusions regarding the determinants of commercial policy.

- Our findings bear out the position NTBs is at least partially governed by economic size, domestic institutions, and the interaction between these factors. More specifically, NTBs are highest in, large states that are characterized by high levels of institutional insulation and autonomy. Thus, states are most likely to impose NTBs when economic incentives to do so exist and when strong domestic institutions insulate policymakers from interest-group pressures, thereby allowing them to advance the national interest unencumbered by those pressure groups that display preferences for freer trade.

- These findings stand in stark contrast to predictions based on either societal or statist models of foreign economic policy artic models—including most endogenous models of protection— emphasize factors related to societal demands for protection, but systematically neglect the factors the regulate the provision of trade barriers.

- Finally, our results yield substantial evidence that tariffs are strongly related to the incidence of NTBs, and that these forms of protection are substitutes. This finding is consistent with the law of constant protection. Among the states considered here, new tariffs could not easily have been imposed due to GATT restrictions. States will low tariff levels that wish to augment their trade barriers therefore have had reason to rely on NTBs for this purpose. Further, states

Notes
characterized by high tariff levels are likely to be sufficiently well-protected that they need not supplement tariffs with NTBs. Our findings suggest the possibility that many of the tariff reductions made by the GATT during the Tokyo Round may not have had the intendeds effect of reducing protection. Instead, these cuts seem to have produced countervailing increases in the incidence of NTBs.

- In addition to the implications of this study, it is also important to address its limitations. One limitation of this sort concern the available data on NTBs. While the coverage ratios on which we relied are viewed by many as the best data of their kind, NTBs are inherently difficult to measure. Further, these data are available for only a few years during the 1980s, and the determinats of NTBs may vary over time. Another limitation concerns the sample of states that was analyzed.

10.6 Key-Words

1. Political economy : The study and use of how economic theory and methods influences political ideology. Political economy is the interplay between economics, law and politics, and how institutions develop in different social and economic systems, such as capitalism, socialism and communism. Political econo*my analyzes how public policy is created and implemented.*

2. *Non-Terrif Barriers* : A form of restrictive trade where barriers to trade are set up and take a form other than a tariff. Nontariff barriers include quotas, levies, embargoes, sanctions and other restrictions, and are frequently used by large and developed economies.

10.7 Review Questions

1. What do you mean by political economy? Discuss.
2. Discuss the societal approach to trade policy.
3. What are the effects of the interaction between societal and statist factors? Explain

Answers: Self-Assessment

1. (*i*) (*a*)   (*ii*) (*d*)   (*iii*) (*b*)   (*iv*) (*c*)   (*v*) (*b*)
   (*vi*) (*a*)   (*vii*) (*b*)

10.8 Further Readings

Books
1. Evans, G., Newnham, J., Dictionary of International Relations; Penguin Books, 1998.
2. Filanlyason, J., Zakher M., The GATT and the regulation of Trade Barriers: Regime Dynamic and Functions; International Organization, Vol. 35, No. 4, 1981.
3. Frieden, J., Lake, D., International political economy: perspectives on global power and wealth, London: Routledge, 1995.
4. Mansfield, E., Busch, M., The political economy of Non-tariff barriers: a cross national analysis; International Organization, Vol. 49, No. 4, 1995.
5. Oatley,T., International political economy: interests and institutions in the global economy; Harlow: Longman, 2007.

Unit 11 : Balance of Payments and Balance of Trade : Meaning and Components

Objectives

After reading this Unit students will be able to:

*   Know the Meaning of Balance of Payments and Balance of Trade.
*   Explain the Components of Balance of Payments.
*   Describe Autonomous and Accommodating Transactions.

Introduction

Balance of payments accounts are an accounting record of all monetary transactions between a country and the rest of the world. By the principles of double entry book-keeping method, monetary receipts are recorded in the credit side of the account and payments in the debit side. The two principal parts of the BOP accounts are the current account and the capital account. The current account shows the net amount a country is earning if it is in surplus, or spending if it is in deficit. The capital account records the net change in ownership of foreign assets. It includes the foreign exchange reserve account, along with loans and investments between the country and the rest of world. Balance of payments in accounting sense should always balance. But that does not imply that the balance of payments will always be in equilibrium — present liabilities are perfectly matched by the present asset positions of the country with respect to the rest of the world. To understand this, one has to understand three basic balances — merchandise trade balance, service trade balance and current account balance.

We often hear that the less developed countries (LDC's) suffer from adverse balance of payments and consequently experience chronic 'foreign exchange gap.' Persistent balance of payments (BOP) deficits have forced countries to resort to corrective measures like currency devaluation, imposition of tariffs, exchange controls, contractionary monetary and fiscal policies and exchange controls of various sorts. Even the so-called developed countries have been no exception to this tendency. Policies of import substitution and export promotion to achieve external balance (or balance of payments equilibrium) have led to serious problems of growth and trade for the countries of the world.

11.1 Meaning of Balance of Payments

The BOP is one of the oldest and the most important statistical statement for any country, especially the more open economies. Put in a nutshell the BOP of any country is "a systematic record of all

economic transactions between the residents of a given country and of the residents of the rest of the world in an accounting period (*viz.* a year)." The system of BOP accounting, some of the concepts and terminologies used in the BOP expression and the interpretation of the BOP categories are of utmost importance to any student of international economics. Despite efforts by international organizations to secure uniformity of classification and presentation, the BOP accounting format differs between different countries. Even the term BOP is somewhat obscure. Yeager, for example, draws attention to the word 'payments' in the term BOP; this gives a false impression that the set of BOP accounts record items which involve only payments. The truth is that the BOP statements record both payments and receipts by a country. It is, as Yeager says, more appropriate to regard the BOP as a 'balance of international transactions' by a country. We will return to this question (and other related questions) later in the chapter. First let us study the system of BOP accounting.

Did u know? The word 'balance' in the term BOP does not imply a situation of comfortable equilibrium; it only means that it is a balance sheet of receipts and payments having an accounting balance.

11.2 Components of Balance of Payments

The BOP transactions include all the foreign receipts of and payments by a country during a given year. The receipts include all the earnings and borrowings of foreign exchange, and they are recorded as credit items. The payments include all the spending and lendings of foreign exchange, and they are recorded as debit items. As such, all the foreign receipts are financial inflows and all the foreign payments are financial outflows in an year. In the purely accounting or book-keeping sense the balance of payments must always balance, because the BOP is a schedule of debit and credit transactions which must necessarily be equal. While the equality of debits and credits (*i.e.* accounting balance) is inevitable, it does not necessarily follow that the BOP equilibrium is guaranteed. Accounting balance is consistent with BOP disequilibrium *i.e.* deficits and surpluses in the BOP.

The BOP statements basically include six major accounts which are as follows :

1. Goods Account
2. Services Account
3. Unilateral Transfers Account
4. Long-term Capital Account
5. Short-term Capital Account
6. International Liquidity Account.

Goods Account

It includes the value of merchandise exports and the value of merchandise imports. These items of foreign exchange earnings and spendings are called as "visible" items in the BOP. If the receipts from exports of goods happen to be equal to the payments for the imports of goods, we describe the situation as one of zero "goods balance."

Otherwise there would be either a positive or a negative goods balance depending on whether we have receipts exceeding payments (positive) or payments exceeding receipts (negative). Positive goods balance is regarded as 'favourable' for a country and negative goods balance is regarded as 'unfavourable.'

Service Account

Just as a country exports goods and imports goods (they may be final consumer goods, intermediate capital goods or raw materials) a country also exports and imports what are called as 'services.' Accordingly, services transactions are regarded as 'invisible' items in the BOP. They are invisible in

the sense that service receipts and payments are not recorded at the port of entry or exist as is the case with the merchandise imports and exports receipts. Except for this, there is no meaningful difference between goods and services receipts and payments. Both constitute earnings or spendings of foreign exchange (as opposed to borrowings and lendings of foreign exchange). Goods and Services Accounts together constitute the largest and economically the most significant components in the BOP of any country.

> *Notes* The *Service Account* records all the service exported and imported by a country in a year. Unlike goods which are tangible or 'visible,' services are intangible.

The service transactions take various forms. They basically include (a) transportation, banking and insurance receipts and payments from and to the foreign countries, (b) tourism, travel services and tourist purchases of goods and services received from foreign visitors to home country and paid out in foreign countries by home country citizens, (c) expenses of students studying abroad and receipts from foreign students studying in the home country, (d) expenses of diplomatic and military personnel stationed overseas as well as the receipts from similar personnel from overseas who are stationed in the home country, and (e) interest, profits, dividends and royalties received from foreign countries and paid out to foreign countries. These items are generally termed as investment income (or expenditure) or receipts and payments arising out of what are called as "capital services". For countries like Malaysia and Singapore which have large foreign investments in their countries or for countries like the USA, the UK, or France and Germany, which have huge investment operation overseas, the investment income payments and receipts constitute a very substantial loss or gain in terms of foreign exchange outflow and inflow. "Service Balance" is the sum of all invisible service receipts and payments, in which the sum could be positive or negative or zero. A positive sum is regarded as favourable to a country and a negative sum is considered as unfavourable. The terms are descriptive as well as prescriptive. Favourable goods and service balance is, therefore, something to strive for and unfavourable goods or service balance is something to avoid, both of which are in the national interest and welfare.

Unilateral Transfer Account

The third account in the BOP schedule is the *Unilateral Transfers Account.* This account includes all gifts, grants and reparation receipts and payments to foreign countries. Unilateral transfer consist of two types of transfers : (a) government transfers and (b) private transfers.

Foreign economic aid or assistance and foreign military aid or assistance received by the home country's government (or given by the home government to foreign governments) constitute government to government transfers. The United States foreign aid to India, for example, is a government transfer constituting a credit item in India's BOP (but a debit item in the US BOP). These are government to government donations or gifts. There is no well worked out theory to explain the behaviour of this account because these flows depend on political and institutional factors. The government donations (or aid or assistance) given to governments of other countries is a mixed bag given for either economic or political or humanitarian reasons. Private transfers, on the other hand, are funds received from or remitted to foreign countries on person-to-person basis. A Malaysian settled in the United States remitting $100 a month to his aged parents in Malaysia, is a unilateral (private) transfer inflow item in the Malaysian BOP. An American pensioner who is settled after retirement in say, Italy, and who is receiving a monthly pension from America is also a private unilateral transfer causing a debit flow in the American BOP but a credit flow in the Italian BOP. Countries that attract retired people from other nations may, therefore, expect to receive an influx of foreign receipts in the form of pension payments. And countries which render foreign economic assistance on a massive scale can expect huge deficits in their unilateral transfer account. Unilateral transfer receipts and payments are also called "unrequited transfers" because as the name itself suggests, the flow is only in one direction with no automatic reverse flow in the other direction.

There is no repayment obligation attached to these transfers because they are not borrowings and lendings but gifts and grants exchanged between government and people in one country with the governments and peoples in the rest of the world.

Long-Term Capital Accounts

It includes the amount of capital that has moved into or out of the country in a year. Any capital that has moved in or out of the country for a period of one year or more is regarded as long-term capital movement. The long-term capital account includes the following categories :

Private direct investment : These investments are done by home country citizens and firms in foreign countries (debit) and by foreigners in the home country (credit). This type of capital movement is induced by differences in profit rate between the home country and the rest of the world.

Private portfolio investment : These investments are done by home country citizens and firms in foreign securities or stocks or bonds or shares (debit) and by foreigners in home country securities, stocks, bonds, shares, etc. (credit). This type of movement in and out of a country is induced by differences in interest rate, dividends or rate of return on capital between the home country's financial assets and those of the foreign nations.

Government loans to foreign governments : These loans are given by home country's government (debit) and to the home government by foreign governments (credit).

If the foreign multinational corporations are investing heavily in our country, we receive capital inflow in the form of direct private investment. It has a favourable effect on our BOP. But when the foreign investors in our country start repatriating profits to their home country, there will be a capital outflow from our country to foreign countries. This goes into our service account as investment income outflow or capital service (debit). When the home country lends out long term capital to foreign countries, we experience capital outflow and hence a debit on our long-term capital account. But when we begin to receive reverse flow in the form of interest on lent out capital or profit on overseas investment the amount will be credited in our BOP as investment income receipt in service account. Capital lending countries would experience deficits on long-term capital account : and capital borrowing countries, like the LDCs, experience surpluses in their long-term capital account.

It is important to note that the long-term capital account includes *new* capital flows into and out of the country; the capital services item in the service account would include the amount of foreign receipts and payments on the cumulative total of *past* long-term capital investments. It is, therefore, possible for a creditor nation that is investing heavily overseas (like the USA) to incur debit or deficits in long-term capital account in the current year while at the same time running capital service item credits or surpluses in service account of equal or even larger magnitude than capital account outflows. A borrowing country, on the other hand receiving credits at present and therefore enjoying surpluses on long-term capital account must soon expect to lose a sizable sum as capital service obligations and, therefore, be ready to suffer deficits on service account. LDCs often experience investment income outflows (capital service account debits) exceeding new long-term capital inflows. The United States, a mature creditor country, regularly earns more from its past investments than it 'loses' in the form of new capital investment outflow. In this sense it is necessary to note that the long-term capital account bears a special relationship to one of the items (investment income item) in the Service Account.

Short-Term Capital Account

The fifth main account in the BOP, is the *Short Term Capital Account.* Bank deposits and other short term payments and credit arrangements fall into this category. Short term capital items fall due on demand or in less than one year, as opposed to long-term capital flows which have maturity after one year or thereafter. The vast majority of short term capital transactions basically represents bank transfers that finance trade and commerce. It is interesting to note that when Malaysian exporter exports rubber worth $5 million to an importer in the United States, it generates a credit of $5 million to the Malaysian merchandise account; but if the US importer pays this sum of $5 million into the

bank account of the Malaysian exporter held in New York bank, the sum of $5 million would be held as debit in Malaysia's short term capital account. The latter constitutes a short term capital outflow of $5 million from Malaysia to the United States.

It is also interesting to note that it is often hard to keep track of all the short term capital movements in and out of the country. They can at best be rough estimates. Indeed in some countries the separate category of short term capital account does not exist. These transactions are simply included in an account under the general term "Errors and Omissions including short-term capital account". This is what is done in the Malaysian system of BOP accounting. In some countries short term capital transactions are included in the "Unrecorded Transactions" as a separate BOP account in its own right. This Unrecorded Transactions Account or Errors and Omissions Account includes, besides short term capital movements, the following items as well.

(a)   Statistical and recording errors

(b)   Smuggling

(c)   Illegal and secret capital movements

(d)   Imperfect estimation procedures.

All of them, like short term capital movements are estimates and are treated as errors and omissions in the BOP accounting. Often it represents a difference in the sums of recorded credit and debit transactions in the first four accounts of BOP (*viz,* goods account, services account, unilateral transfers account and long-term capital account). The fifth account in BOP schedule may therefore be called either as Short Term Capital Account or as Errors and Omissions including short term capital or simply as Unrecorded Transactions Account.

International Liquidity Account

The sixth and final BOP account is the *International Liquidity Account* which simply records net changes in foreign reserves. Essentially this account lists internationally acceptable means of settling international obligations. International Liquidity Account is best understood as follows :

(A)   In the following table total receipts on the first five accounts exceed the total payments on the same five accounts by a sum of $150 million.

Table 1 : Surplus Case ($ Million)

	Credit (Receipts)	Debit (Payments)
1. Goods Account	1,500	800
2. Services Account	500	1,400
3. Unilateral Transfers Accounts	100	120
4. Long Term Capital Account	900	400
5. Errors & Omissions (including short term capital) Account	500	630
6. International Liquidity Account		150
7. Balance of Payments	3,500	3,500

The total receipts are $3,500 million and total payments are $3,350 million. There is a net BOP surplus amounting to $150 million. This sum of $150 million is entered into International Liquidity Account as debit. The logic of accounting for this sum of $150 million as debit or payment is that, this sum represents either

(a)   purchase or import of gold worth $150 million; or

(b)   net addition to accumulation of foreign reserves of $150 million; or

(c)   capital lending in the sum of $150 million to other countries on short term or long term basis.

Notes

The International Liquidity Account in this case represents the BOP surplus magnitude and only shows how the BOP surplus is entered or accounted for in the balance sheet. A debit entry in the International Liquidity Account shows that there is a surplus in the BOP of the country for that year.

(B) The following table has the exact opposite picture. The sum of debit payments ($3,500 million) exceeds the sum of credit receipts ($3,350 million) by $150 million which represents the net deficit in the BOP due to the first five accounts in the table.

Table 2 : Deficit Case ($ Million)

	Credit (Receipts)	Debit (Payments)
1. Goods Account	800	1,500
2. Service Account	1,400	500
3. Unilateral Transfer Account	120	100
4. Long Term Capital Account	400	900
5. Errors & Omissions (including short term capital) Account	630	500
6. International Liquidity Account	150	
7. Balance of Payments	3,500	3,500

The question to ask here is, how was this deficit of $150 million financed ? The answer is that it was financed in one of the following three ways :

(a) selling or exporting gold worth $150 million; or

(b) drawing down upon the past accumulated foreign reserves equal to the sum of $150 million; or

(c) borrowing capital in the sum of $150 million on short term or long term basis from friendly countries or international institutions, like the International Monetary Fund.

The International Liquidity Account in this case, then, represents the BOP deficit sum of $150 million. This amount is entered as credit item to indicate how the sum of $150 million was brought in to finance the deficit of that magnitude arising out of the first five accounts in the BOP schedule. A credit entry in the International Liquidity Account shows, therefore, that the country had a deficit in its BOP of that magnitude in that particular year.

Having understood the six major BOP accounts, it is possible now to study the important concepts and distinctions that one comes across in BOP discussions. Before we do that, let us take a look at the following sample of BOP schedule using some hypothetical numbers in each of the six accounts.

Table 3 : Balance of Payments Schedule—A Sample

Major Accounts	Credit (Receipts)	Debit (Payments)	Net Surplus (+) or Deficit (–)
1. Goods Account	200	180	+ 20
2. Services Account	100	250	– 150
A. BALANCE OF TRADE (1 + 2)	(300)	(430)	(– 130)
3. Unilateral Transfers Account	300	120	+ 180
B. BALANCE OF PAYMENTS ON CURRENT ACCOUNT (1 + 2 + 3)	(600)	(550)	(+ 50)

4.	Long Term Capital Account	150	120	+ 30
C.	**BASIC BALANCE (1 + 2 + 3 + 4)**	**(750)**	**(670)**	**(+ 80)**
5.	Short Term Capital Account	50	40	+ 10
D.	**BALANCE OF PAYMENTS ON CAPITAL ACCOUNT (4 + 5)**	**(200)**	**(160)**	**(+ 40)**
E.	**OVERALL BALANCE OF PAYMENTS (B + D)**	**(800**	**(710)**	**(+ 90)**
6.	International Liquidity Account Net Changes in External Reserves		90	
F.	**BALANCE OF PAYMENTS ACCOUNTING BALANCE**	**(800)**	**(800)**	**(0)**

In Table 3, the six accounts are numbered from 1 to 6, whereas the major BOP concepts are serialized as A, B, C, D, E and F. Using this table as the basis we shall now study the BOP concepts.

11.3 Meaning of Balance of Trade

Balance of trade may be defined as the difference between the value of goods and services sold to foreigners by the residents and firms of the home country and the value of goods and services purchased by them from foreigners. In other words, the difference between the value of goods and services exported and imported by a country is the measure of balance of trade. If the two sums (1) value of exports of goods and services, and (2) value of imports of goods and services are exactly equal to each other, we say that there is balance of trade equilibrium or balance; if the former exceeds the latter, we say that there is balance of trade surplus; and if the latter exceeds the former, then we describe the situation as one of balance of trade deficit. Surplus is regarded as favourable and deficit as unfavourable. In Table 3, there is a balance of trade deficit equal to $130 million.

The readers are warned about the use of terminology. The balance of trade definition adopted above, is that of James E. Meade — a Nobel Prize British economist, W.M. Scammel also prefers to adopt Meade's definition of balance of trade. But some writers however define balance of trade as the difference between the value of merchandise (or goods) exports and the value of merchandise (or goods) imports, making it the same as the 'good balance' or the 'balance of merchandise trade'. There is no doubt that the balance of merchandise trade is of great significance to the exporting country; but of still greater significance is the balance of trade defined in Meade's sense *(i.e.* to lump goods and services balance together). In the familiar macro-economic equation $y = C + I + G + (X − M)$, the expression Net Exports (or X − M) denotes the balance of trade in Meade's sense. Balance of trade is a national income injection and for that reason, it is better to use Meade's concept of balance of trade. Equating balance of trade with goods balance alone is to ignore the importance of service balance as a factor in determining national income.

In case of countries like Malaysia, goods balance is always favourable but service balance is always unfavourable. According to Meade's definition of balance of trade, Malaysia will have a consistent balance of trade deficit; but if we use the other (and more commonly used) definition of balance of trade (synonymous with goods balance alone) then Malaysia's balance of trade will be one of consistent surpluses. Malaysian government itself uses this more commonly used definition of balance of trade and not the one given by Meade. In our Table 3 we have balance of trade surplus of $20 million using the more commonly used definition of trade balance; but if we adopt Meade's definition, there is a balance of trade deficit of $130 million. If any case, before coming to any conclusions on balance of trade surpluses and deficits in a given country, we must first make sure what definition of balance of trade is adopted in that country.

Regardless of which definition is adopted one thing is certain, *viz,* that balance of trade is a national income injection; and hence it is appropriate to regard an active balance *(i.e.* an excess of credits over debits) as a desirable state of affairs. Should this then be taken to imply that a passive trade balance

(*i.e.* an excess of debits over credits) is necessarily a sign of undesirable state of affairs in a country ? The answer is "no". Because, take for example, the case of a developing country, which might be importing vast quantities of capital goods and technology to build a strong agricultural or industrial base. Such a country in the course of doing that might be forced to experience passive (or adverse) trade balance, and such a situation of passive trade balance cannot be described as one of undesirable state of affairs. This would therefore again suggest that before drawing meaningful inferences as to whether passive trade balance is a desirable or an undesirable state of affairs for a country, we must also know the composition of imports which are causing conditions of adverse trade balance for that country. At any rate, the importance of trade balance in any country's BOP can hardly be exaggerated.

Balance of Payments on Current Account

This is a broader concept than the concept of balance of trade. Balance of Payments on Current Account includes the sum of three balances *viz.* Merchandise balance, service balance and unilateral transfers balance. In other words, it comprises of trade balance (in Meade's sense) and transfers balance. In Table 3 the positive unilateral transfers balance of $180 million is added on to the negative trade balance of $130 million which will give us a current account BOP surplus of $50 million.

Balance of Payments on Current Account is also referred to as Net Foreign Investment because the sum represents the contribution of foreign trade to GNP.

It is also worth remembering that BOP on Current Account covers all the receipts on account of earnings (or opposed to borrowings) and all the payments arising out of spendings (as opposed to lendings). There is no reverse flow entailed in the BOP current account transactions. This is in sharp contrast to the balance of payments on Capital Account which we will see below.

Balance of Payments on Capital Account

For a long time, economists had assumed that factors of production do not move across international boundaries; the classical economists built models of trade assuming that only goods and services move across international boundaries. International capital movements viewed in that light, were an impossibility. Perhaps, for this reason, we do not have a well developed theory of international capital movement, although, as we have seen in previous chapters, we have well advanced theories of international trade in goods and services (*e.g.* Comparative Advantage Models, Heckscher-Ohlin model, demand reversal, factor-intensity reversal, Leontief Paradox and a whole lot of the so-called "New" theories of international trade). However, this is no occasion to discuss the theory of international capital flows. Theory or no theory, international capital movements in and out of countries are a fact of life and very much a reality in today's world. With so many multinational banks, transnational corporations with their giant global operations, inter-governmental aid, grants and loans and international institutional arrangements for borrowing and lending of money between the countries of the world, the international capital and investment flows across nations have reached unprecedented proportions, especially after World War II period. Less developed countries are the net recipients of foreign capital and investment and some see it as an opportunity for these countries to maximize their rate of growth and minimize their balance of payments hardships. The radical political economists consider this trend to be potentially dangerous to the LDCs, because (a) it will subject the economies of the poor Third World nations to economic imperialism by the Western Capitalist countries and (b) the mounting foreign debt of these Third World countries keeps compelling them to borrow still more in order to repay the debt resulting not in less debt but more debt as the years go by. They will be, therefore, perpetually "dependent" on borrowings from foreign sources. Be that as it may!!

Returning to the question of BOP accounting procedure, all the transactions involving inward or outward movement of capital and investment (be it long term or short term, direct or portfolio, private or government, individual or institutional, tied or untied, interest bearing or non-interest bearing, soft or hard) are included in the Capital Account of the BOP of the reporting country. In simple terms, the BOP Capital Account comprises of the Long-Term and Short-Term Capital Accounts. In Table 3, the Capital Account balance shows a net surplus of $40 million (see item numbers 4 and 5 and item D in the Table).

It is useful to understand the broad trends and implications of Capital Account transactions in the BOP of countries of the World Developed Countries (DCs) are the net exporters of capital and investment and the less developed countries (LDCs) are the net borrowers of foreign capital and investment. From that it follows obviously that the DCs would experience deficits (or adverse trends) in their BOP capital accounts; the LDCs, on the other hand, would "enjoy" capital account surpluses in their BOP. The questions that arise here are : Is this capital account surplus good ? And is this capital account deficit bad for the countries ? The answers are somewhat as follows :

LDCs are net borrowers of foreign capital and recipients of foreign investment, and to that extent they would "enjoy" favourable BOP trends. This is undoubtedly true. But sooner or later this foreign capital and investment will leave the LDCs and go back. Whether they do or do not go back to their home country, what is most certainly true is, that the returns on that capital and investment in the form of profits, interest, dividends and royalties would be repatriated from the "host" countries to the "home" countries (in this case from LDCs to DCs). And this sum would create deficit tendencies in the Current Account of the BOP of the LDCs concerned. In other words, capital account surplus of the present year will create current account deficits of a potential nature (in the form of investment income outflows) for the years ahead. In that sense, the country which "enjoys' a capital account surplus today must get ready to "suffer" a current account deficit in future. The prognosis is entirely correct but it may not result in a nightmare. Because by making productive use of foreign capital and investment and increasing both the GNP and export capacity the LDC can avoid future BOP deficits on current account *i.e.* they can offset investment income outflows and capital repatriation by increasing merchandise exports as well as service exports. It does not, therefore, necessarily follow that a capital account surplus today is a sure sign of current account deficit tomorrow (or in 10 years from now). It much depends on the manner in which foreign capital and investment are put to use in the receiving country. If they are put to unproductive use resulting in no expansion of real output of goods and services, then of course, it would be true to say that today's capital account surplus is tomorrow's current account deficit.

The countries which "suffer" capital account deficit today need not worry because it will automatically result in BOP current account surpluses (arising out of investment income inflows) at a future date. Today's deficit (in capital account) is truely tomorrow's surplus (in current account). There can be little doubt in saying so. Only in the unlikely events where no investment income inflow ever materialized (due to losses on investment) or permitted to materialize (due to host country government's radical policies like expropriation, nationalization etc.) would it be possible to argue that today's capital account surplus is no guarantee of tomorrow's current account surplus?

However, the significance of BOP deficits and surpluses arising out of transactions in capital account can, therefore, be seen only with a time perspective and future prospects clearly in mind. Only then, can the significance of capital account in the BOP be fully understood.

Basic Balance

This is a relatively straightforward and simple concept. Basic balance in the BOP comprises of the BOP on current account plus long-term capital account. The short term capital account balance is not included in the basic balance. This is perhaps for two main reasons — (a) short term capital movements, unlike long-term capital flows, are relatively volatile and unpredictable. They move in and out of a country in a period of less than a year or even sooner than that. It would, therefore, be improper to treat short term capital movements on the same footing as current account BOP transactions which are extremely durable in nature. Long-term capital flows are relatively more durable and, therefore, they qualify to be treated along side the current account transactions to constitute basic balance, (b) in many cases countries do not have a separate short term capital account for reasons discussed earlier in this chapter; in these countries, short term capital transactions constitute a part of the "Errors & Omissions Account". Hence the justification in excluding short term capital flows from the definition of "Basic Balance".

An active basic balance is a good sign and a passive basic balance is a bad sign for the reporting country's overall BOP picture.

Overall Balance of Payments

This is a sum of balance on Current Account and on Capital Account put together. It includes all international monetary transactions of the reporting country *vis-a-vis* the rest of the world. It is highly aggregative, and like any other highly aggregated variables, the concept (or the sums entered as credit and debit under this item) cannot be of much significance. Because, the aggregate credit and debit figures do not reveal the behaviour of change of the components which constitute the aggregate. Take for instance, the question of surplus and deficit. Can we say that a surplus in the overall balance of payments is a good sign *or* a deficit in the overall balance of payments is a bad sign ? We cannot give sensible answers to these questions simply by reading the entries in the credit and debit columns of the overall BOP row. We must not only know the *extent* of BOP overall surplus or deficit but also the *location* of those surpluses or deficits. The answers can then be given something like these :

(i)   If the overall surplus in the BOP was caused by current account surpluses but not capital account surpluses, then the surplus may be a good sign for the country.

(ii)   If the overall deficit in the BOP was caused by current account deficits rather than capital account deficits, then the deficit may be considered as a bad sign for the reporting country.

In other words, if there is an overall surplus, we will have to first locate whether the surplus originated in current account or capital account or both. The same will have to be done in case of a deficit in the overall BOP. Therefore, the overall BOP figures by themselves, whether they indicate a surplus or a deficit, do not reveal the real situation. For this reason not much economic significance can be attached to the overall BOP concept. The current and capital account breakdown is very useful and significant.

Accounting Balance of Payments

The overall BOP entries (item E*)* in our Table 3 show a net surplus of $90 million (total credits of $800 million *minus* total debits of $710 million arising out of all the transactions in the five BOP accounts above). This sum of $90 million surplus is entered into the International Liquidity (debit) Account. The rationale behind this entry in the debit column is that, this sum of $90 million constitutes disposal of that BOP surplus in any of the following ways (a) purchase or import of gold worth $90 million; or (b) adding to the country's stock of foreign exchange reserves of $90 million for future use; or (c) extending short term loan of $90 million to other needy countries or buying some foreign income-earning short term assets. There may even be some combination of (a), (b) and (c) as well. In any case International Liquidity Account debit entry indicates a surplus in the overall BOP and the way it is settled in or the other of the three methods states just above.

Having done that in the International Liquidity Account (item 6 in Table 3), what we then notice in the Accounting BOP column (item F in Table 3) is an exact equality in the BOP debit and credit entries to give us an accounting balance. Notice that the balance of payments must always balance in the accounting or book-keeping sense. This is because for any surplus (or deficit) in the overall BOP (item E in Table 3) there must be a corresponding debit (or credit) entry of the equivalent amount in the International Liquidity Account such that the total credits and debits in the Accounting BOP item must show an exact equality. Regardless of whether the country has an actual surplus or a deficit in its overall BOP, there can be no surplus or a deficit in any country's balance of payments in the accounting sense. In other words, the balance of payments in the accounting sense is consistent with the surplus or deficit which the country may have. International Liquidity Account cancels all the surplus or the deficit in the BOP to give a zero accounting surplus or the deficit. Since "balance of payments must always balance" there can, therefore, be no 'imbalance' in the BOP. Then what is the meaning of 'deficit' and 'surplus' in balance of payments ? If there can be no 'imbalance' then what is the meaning of 'disequilibrium' in the BOP ? Why do countries talk and worry about BOP 'adjustment' and BOP 'settlement' ? Now we will turn our attention to these and other related questions and distinctions.

11.4 Autonomous and Accommodating Transactions

A distinction is made between the autonomous and the accommodating *or* above the Line and below the Line transactions in the BOP. The distinction is useful to define the concepts of 'deficit' and 'surplus' in the BOP.

Autonomous or above the Line transactions are those that "take place regardless of the size of other items in the balance of payments". Take, for example, the export of goods to a foreign buyer. It is an 'initiating' or an 'autonomous' transaction and its value results in payments by foreigners to the home country, which is entered as a credit item. When the home country borrows $100 million from the World Bank to construct a highway, the sum of $100 million is credited into the Long Term Capital Account of the home country. If the foreign multinational corporations repatriate $300 million of their profits (earned from their investment operations in our country) to their country of origin, then we enter $300 million into our BOP Service Account as investment income outflow or capital service debit item. If India receives $50 million from the United States as aid for the famine-stricken people in India, then the amount of $50 million is entered as credit into India's Unilateral Transfer BOP Account. All these transactions which take place either in the goods account, or the services account, or the unilateral transfer account or the long term and short term capital account of a country are considered as 'autonomous' or 'above the line' transactions. They arise out of autonomous economic activities as credit or debit transactions, and these transactions take place regardless or independent of balance of payments situation.

The 'accommodating' or 'below the line' transactions, on the other hand, take place on account of, or due to, balance of payments situation of a country. They are the *result* of balance of payments situation, whereas the 'autonomous' transactions are the *cause* of balance of payments situation. Take for example, gold exports or foreign borrowings. Suppose South Africa, a gold exporting country, exports $800 million worth of gold as a commodity export, then these $800 million export proceeds are entered as credit in that country's merchandise account. Here gold is exported as an autonomous activity or as a current account transaction. This activity *causes* foreign exchange earnings and thereby *determines* the BOP situation for the country (note that it is not determined by the BOP situation of the country).

On the other hand, if India (which is not a gold exporting country) is forced to export gold worth $400 million to settle its balance of payments deficits, then we say that this gold export is not an 'autonomous' but an 'accommodating' transaction undertaken exclusively with a view to solve its BOP problem. Here, the gold export is the result of BOP situation and not its cause. This 'accommodating' gold export transaction is entered into India's BOP International Liquidity Account as a credit item. Another example would be, where a country has borrowed from abroad. If Malaysia has borrowed $700 million with a view to construct its East West Highway from the World Bank, it is an 'autonomous' transaction, and it is entered into Malaysia Long Term Capital Account as a credit entry. But if Malaysia has borrowed a sum of $700 million from the World agencies to settle its BOP deficit, then it is treated as an 'accommodating' transaction and 'entered into the country's International Liquidity Account as a credit item.

In brief, all credit and debit entries in the BOP current and capital accounts are regarded as 'autonomous' or 'above the line' transactions; and all the credit and debit entries in the International Liquidity Account are to be regarded as 'accommodating' or 'below the line transactions'. The distinction between 'autonomous' and 'accommodating' transactions lies in the questions whether the transactions have *caused* the BOP situation or whether the transaction has been *caused by* the BOP situation.

The distinction between the 'autonomous' and 'accommodating' transactions looks very clear cut from what has been said above. But the distinction may sometimes be not all that cut and dry. For instance, if Malaysia exports $500 million worth of rubber, we take it as an 'autonomous' export transaction and treat it as now having been caused by the country's BOP situation. But it may be possible that Malaysia exported this $500 million worth of rubber to *avoid causing* a BOP problem. In this case, this transaction has been undertaken *due to* BOP considerations and not *regardless* of BOP considerations. The subtlity lies in the question whether this export transaction was undertaken regardless of BOP considerations or whether it was undertaken with BOP considerations in mind *(i.e.,* to avoid a possible BOP deficit problem). This raises the question of short term and long term motives behind all BOP transactions of a country, and the problems of identifying transactions as 'autonomous' or 'accommodating' become extremely hard to handle. For accounting purposes, it is reasonable to treat all current and capital account transactions as 'autonomous' or 'above the line'; and all the entries in the International Liquidity Account as 'accommodating' or 'below the line'

transactions. Accommodating transactions are undertaken for the deliberate (or explicit) purpose of financing any imbalance in autonomous transactions.

Now, we are in a position to define BOP 'deficit' and 'surplus'. A 'deficit' in the BOP occurs when the autonomous payments (debits) exceed the value of autonomous receipts (credits); and a 'surplus' results when the autonomous credit receipts exceed the autonomous debit payments. If the two sums, autonomous receipts and payments, are equal, we have an 'equilibrium' in the BOP. Alternatively, the sum of accommodating transaction entered as credit in the International Liquidity Account is the measure of 'deficit' in the BOP, while a similar entry in the debit column of that account, is the measure of 'surplus' in the BOP. If, however, the International Liquidity Account shows entry of a zero sum as credit or as debit, then there would be 'equilibrium' in the BOP of a country. A 'deficit' in the BOP is considered, as negative or unfavourable or adverse BOP situation for the reporting country; and a 'surplus' in the BOP is considered as positive, favourable or active BOP situation of the country. BOP 'equilibrium' is considered as the desirable external economic state of affairs for any country. It should be remembered that 'balance' (or imbalance') in the BOP is different from 'equilibrium' (or 'disequilibrium') in the BOP of a country. Strictly speaking, 'balance' is an accounting balance and it must be there all the time for any country; 'equilibrium', on the other hand, can be there only when the autonomous credit receipts equal autonomous debit payments. 'Equilibrium' is an economically meaningful concept, whereas 'balance' is only an accounting or a book keeping term. In common parlance the word BOP 'balance' is used interchangeably with the term BOP 'equilibrium', but the difference between the two terms must be borne in mind. Similarly, the 'imbalance' is used interchangeably with 'disequilibrium' in the BOP. Here again, strictly speaking, there can never be an 'imbalance' in the accounting or book-keeping sense. 'Disequilibrium' is possible when there is either a BOP deficit or a surplus. The more meaningful expressions would be BOP 'equilibrium' or 'disequilibrium' rather than BOP 'balance' or 'imbalance'.

This brings us now to the question whether 'deficit' in the BOP is always bad and a 'surplus' in the BOP always good as it is commonly suggested. While treating 'disequilibrium' in BOP as undesirable, we generally tend to associate 'disequilibrium' only with a 'deficit'. We, generally, consider BOP 'surplus' as good and, therefore, we do not identify 'surplus' with 'disequilibrium' *i.e.* something that is as undesirable as deficit. As a matter of fact it can be argued that if 'surplus' is good, then 'deficit' could also be good in certain circumstances; and if 'deficit' is bad, then 'surplus' could as well be bad for many reasons. The full significance of all these can be understood only when we understand the two aspects of the BOP disequilibrium : *viz.* its *location,* and its *duration.*

First, consider the aspect of *location.* A surplus in the current account should be treated as generally favourable but a surplus in the capital account may not necessarily be a good sign, because the former *(i.e.* current account surplus) is a sign of earning capacity whereas the latter *(i.e.* capital account surplus) only indicates a capacity to borrow. Borrowings, unlike earnings, entail commitments to repay at some future date. We also know that it is better to earn rather than borrow and 'enjoy' a current account surplus rather than a capital account surplus. Under some circumstances borrowing may be good as well. If, for instance, a country has borrowed $100 million to invest in some useful project, then such a borrowing is productive, because it will result sooner or later in both the GNP increase and the increase in repaying capacity. The borrowing in that case has been self-liquidating and GNP increasing as well. Viewed in this sense, BOP capital account surpluses may be good. It would be tedious and unnecessary to spell out the practical implications of having BOP capital account surpluses and using that surplus for unproductive consumption (or investment) which results neither in the GNP expansion nor in repaying capacity for the country.

Similarly, for the 'deficit' case. A current account deficit indicates spendings in excess of earnings with its obvious implications, while a capital account deficit would mean that the nation is undertaking net foreign lendings. A mature creditor country like the United States would be incurring such capital account deficits in BOP. Its effects on the US *present* BOP situation may be unfavourable; but sooner or later, when the money that is lent out begins to come back together with the returns on that foreign lending in the form of interest, profits, dividends etc., then it will produce favourable effects on the US BOP in *future.*

Boardly, certain BOP principles can be laid down with global perspectives in mind (a) current and capital accounts with similar signs are both undesirable; they tend to aggravate problems of international liquidity and hamper growth of world trade, (b) the correct behaviour for a country with a surplus on current account is to lend abroad and run deficits on capital account.

The second aspect of BOP deficit or surplus is its *duration.* If the deficit in the BOP, regardless of its location, is of a temporary nature, say on account of export crop failure, or political events which might have led to temporary capital flight, then such a transient effect on the BOP may not be a cause of serious concern. If the BOP deficit is fundamental or persistent, especially arising out of current disequilibrium, such a prolonged deficit is bad. It is suggestive of some structural disequilibrium which calls for corrective action. Current account deficits of short term duration can be handled by short term or long term foreign borrowings. But the current account deficits which are repeated year after year cannot and should not be financed by foreign borrowings alone. In the final analysis, exports have to pay for imports and it would be necessary to reverse the current account surpluses, like persistent current account deficits, need to be given close scrutiny, because over the years they increase foreign indebtedness and make the country's BOP situation extremely vulnerable. In short, if the deficits or surpluses in current or capital account are of a temporary nature, they do not constitute a problem; but if they become persistent, they need corrective policy action.

Now let us come to the normative judgement issues on 'deficit' and 'surplus' questions in current account. Generally it is believed that current account deficits are bad and surpluses are good. It is easy to see how deficits in current account are bad, especially when the deficits are prolonged and lingering. What about current account surpluses ? Are they unquestioningly good for a country ? Here again one may readily agree that occasional or temporary surpluses are good, and especially so, if these surpluses are essential (a) to offset capital account deficits or (b) to offset investment income outflows or to pay off old debts. If, on the other hand, a country is running persistent surpluses in its current account, not offset by capital account deficits, it will run several risks : (a) accumulation of foreign (official) reserves by the country's monetary authority; which leads to expansion in money supply and domestic inflationary price upsurge or (b) exchange rate appreciation which would in turn make the country's exports less competitive in the world market; which leads to import expansion and export contraction. The precise nature of the result depends of course on the country's exchange rate policy, monetary and fiscal policy and exchange controls policy. In any case, persistent surpluses in the BOP do not necessarily produce healthy effects on the economy. It is necessary to realize that if BOP deficits are not good, so is the case with BOP surpluses. BOP 'disequilibrium' denoting as it does, something undesirable, applies equally to 'deficits' and 'surpluses' alike, although deficits pose a more serious threat and a more difficult challenge than the BOP surpluses.

Balance of Payments Settlement and Adjustment

A distinction is made between BOP *settlement* and BOP *adjustment* and this distinction can be understood if we proceed as follows. Suppose, we have a deficit (or a surplus) in the current and capital accounts of our country's BOP, then we can "settle" this deficit (or surplus) arising out of imbalance in the autonomous transactions by accommodating payments (or accommodating receipts in case of a surplus) so as to produce a BOP balance in the accounting sense. In this case, therefore, what we have undertaken, is a process of BOP *settlement.* We have, in this case, merely and temporarily overcome the BOP problem with the help of necessary accommodating transactions. The fire, so to speak, has been extinguished but the house is still not reconstructed. If, on the other hand, we control the deficit (or surplus) by controlling the forces which were causing this deficit (or surplus) then we have undertaken what may be called as BOP *adjustment.* BOP adjustment is said to have taken place only when we have produced 'balance' in autonomous transactions *i.e.* when autonomous credit receipts are equal to autonomous debit payments. In brief, when accounting balance is produced with the help of accommodating transactions there is said to be BOP settlement; and when this balance is produced without the help of accommodating transactions, there is said to be BOP adjustment. Adjustment is more desirable and more difficult than settlement.

'Full Employment' Equilibrium or 'True' Balance

Finally, a word on what is called 'full employment' equilibrium or 'true' balance in BOP. We have already seen that when the sum of accommodating transactions is zero, there is 'equilibrium' in the balance of payments (as well as a 'balance' in BOP). It is, however, possible that such an 'equilibrium' has been produced by (a) imposing trade and payment restrictions such as import tariffs, import quotas, export duties, restrictions on foreign travel, exchange controls, exchange rate support policies and other monetary and fiscal policy applications; and (b) by causing internal imbalance *i.e.* by causing inflation, or unemployment in the economy. If the BOP 'equilibrium' is produced by causing (a) or (b) or both, then it is not to be considered as 'true' balance or 'full employment' equilibrium.

If, on the other hand, equilibrium in the BOP is produced without (a) using commercial policy and (b) causing inflationary or deflationary gaps in the GNP of the country, then what we have is a 'true' balance or 'full employment' equilibrium in the BOP. As such, 'full employment' equilibrium or 'true' balance is inconsistent with the existence of commercial policy and internal imbalance (defined as presence of inflation or unemployment). Ragnar Nurkse and James E. Meade, have introduced this concept of 'full employment' equilibrium or 'true' balance. In their terminology, 'external balance' must be consistent with 'internal balance' on the one hand and 'free trade' in goods and services and factor flows on the other.

Balance of Payments and Economic Policy

Before going into this question, a word of caution must be inserted at the very outset. National accounting system is not completely standardized among countries; different countries use different systems. This is reflected also in the construction of balance of payments. There is agreement on the broad features, but differences in detail do exist. Our aim is to understand how basic questions of economic analysis and policy can be elucidated by the use of BOP statistics and to gauge what factors one has to take into account to be able to interpret a country's BOP situation in any meaningful sense.

Having understood the sense in which the BOP is always in 'balance' and the sense in which the BOP might be in 'disequilibrium' one can talk about suggested economic policy in relation to a given BOP situation of a country. For example, if a country has a deficit and an accommodating capital inflow, it must in general try to implement policy measures aimed at reducing the deficit; but a country with a surplus in its BOP and an accommodating capital outflow need not take immediate measures. Because surpluses do not usually create great problems, so we are not specially concerned with surplus countries. For economic policy purposes one is specially concerned with BOP problems of deficit countries.

To understand the nature of a deficit, one has to judge it against the background of the general economic policy of a country and the policy options, the country has at its disposal. If a country is already pursuing a tight monetary and fiscal policy and has tariffs and import controls, but yet it has a serious deficit, it may be very difficult for such a country to get rid of a deficit. We can then talk about *actual* and *potential* deficits. The actual deficit which has appeared on the surface is in that case, much smaller than the potential deficit that could have surfaced but has indeed been suppressed by tight domestic and foreign trade economic policies of the country. The possibility or scope of pursuing a more restrictive policy to close the actual deficit may no longer exist for a country, because it already has reached its upper limit. Furthermore, if the economy is already experiencing politically unacceptable levels of high unemployment, it will be almost impossible to try to cut down BOP deficits by pursuing still contractionary monetary, fiscal and other policies. In such tight situations only international capital flows can play a vital role in equilibriating the BOP. Once again the nature of capital flows is very crucial. We have already said that accommodating capital inflows, especially, if they are continued over several years, are a sure warning signal. It is left to the ingenuity of the country' planers and policy makers to adopt ways and means of converting accommodating capital imports of short-term nature into planned long term autonomous capital imports. If that can be done, the country need not change its economic policy. It can proceed without having to worry about the BOP situation for the next 15-20 years. We shall return to these policy-oriented questions concerning BOP later in the other chapters; but in the following chapters let us study the relation between foreign trade and the national economy.

Self- Assessment

1. **Choose the correct option:**

    (*i*) A nation's balance of trade is equal to its exports less its imports of

    (*a*) goods   (*b*) goods and services

    (*c*) financial assets   (*d*) official reserves

    (*ii*) A nation's balance on the current account is equal to its exports less its imports of

    (*a*) goods and services

    (*b*) goods and services, plus Canadian purchases of assets abroad

    (*c*) goods and services, plus net investment income and net transfers

    (*d*) goods and services, minus foreign purchases of assets in Canada

    (*iii*) The net investment income of Canada in its international balance of payment is the

    (*a*) interest income it receives from foreign residents

    (*b*) dividends it receives from foreign residents

    (*c*) excess of interest and dividends it receives from foreign residents over what it paid to them

    (*d*) excess of public and private transfer payments it receives from foreign residents over what it paid to them

    (*iv*) A nation may be able to correct or eliminate a persistent (long-term) balance of payments deficit by

    (*a*) lowering the barriers on imported goods

    (*b*) reducing the international value of its currency

    (*c*) expanding its national income

    (*d*) reducing its official reserves

    (*v*) If exchange rates float freely, the exchange rate for any currency is determined by the

    (*a*) demand for it

    (*b*) supply of it

    (*c*) demand for and the supply of it

    (*d*) official reserves that back it

    (*vi*) If a nation had a balance of payments surplus and exchange rates floated freely, the foreign exchange rate for its currency would

    (*a*) rise, its exports would increase, and its imports would decrease

    (*b*) rise, its exports would decrease, and its imports would increase

    (*c*) fall, its exports would increase, and its imports would decrease

    (*d*) fall, its exports would decrease, and its imports would increase

11.5 Summary

*   The BOP is one of the oldest and the most important statistical statement for any country, especially the more open economies. Put in a nutshell the BOP of any country is "a systematic record of all economic transactions between the residents of a given country and of the residents of the rest of the world in an accounting period (*viz.* a year)." The system of BOP accounting, some of the concepts and terminologies used in the BOP expression and the interpretation of the BOP categories are of utmost importance to any student of international economics.

*   The BOP transactions include all the foreign receipts of and payments by a country during a given year. The receipts include all the earnings and borrowings of foreign exchange, and they are recorded as credit items. The payments include all the spending and lendings of foreign

exchange, and they are recorded as debit items.

- The sixth and final BOP account is the *International Liquidity Account* which simply records net changes in foreign reserves. Essentially this account lists internationally acceptable means of settling international obligations.

- Balance of trade may be defined as the difference between the value of goods and services sold to foreigners by the residents and firms of the home country and the value of goods and services purchased by them from foreigners. In other words, the difference between the value of goods and services exported and imported by a country is the measure of balance of trade. If the two sums (1) value of exports of goods and services, and (2) value of imports of goods and services are exactly equal to each other, we say that there is balance of trade equilibrium or balance; if the former exceeds the latter, we say that there is balance of trade surplus; and if the latter exceeds the former, then we describe the situation as one of balance of trade deficit. Surplus is regarded as favourable and deficit as unfavourable. In Table 3, there is a balance of trade deficit equal to $130 million.

- For a long time, economists had assumed that factors of production do not move across international boundaries; the classical economists built models of trade assuming that only goods and services move across international boundaries. International capital movements viewed in that light, were an impossibility.

- This is a sum of balance on Current Account and on Capital Account put together. It includes all international monetary transactions of the reporting country *vis-a-vis* the rest of the world. It is highly aggregative, and like any other highly aggregated variables, the concept (or the sums entered as credit and debit under this item) cannot be of much significance.

- A distinction is made between the autonomous and the accommodating *or* above the Line and below the Line transactions in the BOP. The distinction is useful to define the concepts of 'deficit' and 'surplus' in the BOP.

- Autonomous or above the Line transactions are those that "take place regardless of the size of other items in the balance of payments". Take, for example, the export of goods to a foreign buyer. It is an 'initiating' or an 'autonomous' transaction and its value results in payments by foreigners to the home country, which is entered as a credit item.

- The distinction between the 'autonomous' and 'accommodating' transactions looks very clear cut from what has been said above. But the distinction may sometimes be not all that cut and dry. For instance, if Malaysia exports $500 million worth of rubber, we take it as an 'autonomous' export transaction and treat it as now having been caused by the country's BOP situation.

11.6 Key-Words

1. Balance of Payment : A record of all transactions made between one particular country and all other countries during a specified period of time. BOP compares the dollar difference of the amount of exports and imports, including all financial exports and imports. A negative balance of payments means that more money is flowing out of the country than coming in, and vice versa.

2. Unilateral transfer : An economic transactions between residents of two nations over a stipulated period of time, usually a calendar year. Typically, these transactions consist of gift exchanges, pension payments and the like, but they can encompass other goods and services as well.

11.7 Review Questions

1. What is the balance of payment? Discuss.
2. Discuss the components of Balance of payments .
3. Distinguish between long- term and short -term account.
4. Explain the meaning of balance of trade.

Answers: Self-Assessment

1. (*i*) (*a*)   (*ii*) (*c*)   (*iii*) (*c*)   (*iv*) (*b*)   (*v*) (*c*)

    (*vi*) (*b*)

11.8 Further Readings

Books

1. Adapted, with slight modifications, from the definition given in the International Monetary Fund *Balance of Payments Manual,* 2nd edn., January 1950, *p.* 1.

2. Leland B. Yeager, *International Monetary Relations* (New York : Harper & Row, 1966), *p.* 36.

3. J.E. Meade, *The Theory of International Economic Policy, Vol. I : The Balance of Payments,* (1951) *p.* 7. See also Yeager, *op. cit., p.* 452.

4. W.M. Scammel, *International Trade and Payments* (Toronto : The Macmillan Company, 1974) *p.* 272.

5. J.E. Meade, *op. cit., p.* 11.

Unit 12 : Equilibrium and Disequilibrium in BOP

CONTENTS

Objectives

Introduction

12.1 Equilibrium of Balance of Payments

12.2 Disequilibrium in Balance of Payments

12.3 Summary

12.4 Key-Words

12.5 Review Questions

12.6 Further Readings

Objectives

After reading this Unit students will be able to:

•   Explain Equilibrium of Balance of Payments.

•   Describe Disequilibrium in Balance of Payments.

Introduction

The main purpose of this unit is to arrive at definitions of balance of payments equilibrium sufficiently precise to furnish policy-makers with meaningful guides. Recent Australian and New Zealand experience, of relaxations in exchange controls followed by tighter controls and of experimental monetary techniques, suggest an imperfect understanding not only of the concepts of equilibrium and reserve adequacy but also of measures most suited to attain these ends. A restatement of the factors involved in defining these concepts together with an appraisal of remedial measures for a balance of payments disequilibrium is considered appropriate.

Disequilibrium in the balance, of payments can arise due to persistently one sided movement of one or more than one trading terms. If, for instance, the total value of goods exported exceeds the total value of the goods imported over a given period and this surplus is not offset by the debit balance on invisible item, the country will have favorable balance of payments. Disequilibrium in the balance arises when exports of a country fall short of imports because of decrease in production at home, due to stiffer competition abroad or of an appreciation in the currency or fall of purchasing power of the buyers in the foreign market.

When the imports remain unaffected or increase, then the country will also face deficit in her balance on invisible items, the country will have disequilibrium in her balance of payments. Disequilibrium in her balance of payments can also arise over a given period due to excessive imports not equalized by exports of invisible items and if it is not offset by credit balance on visible items, the country will face disequilibrium in her balance of payments.

12.1 Equilibrium of Balance of Payments

Definition and Explanation

"The **equilibrium of balance of international payment** is a statement that takes into account the debits and credits of a country on international account during a calendar year".

When a country has unfavorable or adverse balance of payments, it is regarded as herald of disaster because the country by having deficit in her balance of payments either decreases her balances abroad or increases her foreign debts. When it has favorable credit balance, it is considered that the country

is heading towards prosperity because by having surpluses, it either increases her foreign credits or reduces her foreign debits.

There is no doubt that a study of country's balance of payment reveals much information about its economic position and development of the country. But when we are to see that a country is heading towards financial bankruptcy or higher standard of living, we are to examine the balance of payments of many years of that country.

A persistent deficit in the balance of payments on current account certainly leads to economic and financial bankruptcy. A continued favorable balance on current account is also disadvantageous because it creates difficulties for other countries. The credit country may utilize her surplus in advancing short or long term loans to the debtor country. But if it gives no opportunity to the debtor country to repay the loan by exporting more, then how can the loans he realized ?

The hard earned surplus of the credit country will then one day be turned into gifts and this may create political difficulties for the creditor country. We have seen, thus that a country should neither have unfavorable nor favorable balance of payment on current account in perpetuity. It must obtain equilibrium in her balance of payments over a reasonable period of time. From this it may not be concluded that a country should balance her account every year with every country with which it has trade relations.

Balance of payments equilibrium occurs when induced balance of payments transactions---those engineered by the government to influence the nominal exchange rate---are zero. This implies that autonomous receipts from exports and the sale of securities abroad equal autonomous payments for imports and the purchase of securities from foreign residents. Since changes in the stock of official reserves of foreign exchange are the method used by the authorities to fix or otherwise manipulate the exchange rate, balance of payments equilibrium requires that the stock of foreign exchange reserves be constant.

Induced transactions are frequent when the exchange rate is fixed---only by chance will autonomous receipts and payments balance. They can also occur when the exchange rate is flexible and the authorities want to influence its movement. But we will concentrate primarily on the fixed exchange rate case here.

> *Notes*   A country may have favorable balance of payment with one country and unfavorable with another but in the long run it must balance her account. The total liabilities and total assets of all nations related to one currency block must balance over a reasonable period of time.

The condition of asset equilibrium---the LM equation---can be presented as the equality of the demand and supply of nominal money balances as follows :

1. $M = mm\,H = mm\,(R + Dsc) = P\left[\gamma - \theta(r^* + \tau) + \varepsilon Y\right]$

   where Y is real income, P is the price level, r^* is the real interest rate determined by conditions in the rest of the world, τ is the expected rate of inflation, mm is the money multiplier, R is the stock of official foreign exchange reserves, which can be thought of as the foreign source component of the stock of base money and Dsc is the domestic source component of the stock of base money H. The above equation can be manipulated to present the equilibrium stock of high powered or base money as

2. $H = R + Dsc = (P\,/\,mm)\left[\gamma - \theta(r^* + \tau) + \varepsilon Y\right]$

   which can be further rearranged to move the stock of foreign exchange reserves R to the left side.

3. $R = (1 / mm) \ P[\gamma - \theta(r* + \tau) + \varepsilon Y] - Dsc$

An increase in the demand for money, given by the P times the expression in the square brackets [..], leads to an increase in the stock of official foreign exchange reserves as the authorities act to maintain the fixed exchange rate. A reduction in the money multiplier increases the stock of base money required to support the existing quantity of money, requiring in an increase in the stock of reserves to provide the additional base money demanded. An increase in Dsc leads to an equal decline in the stock of foreign exchange reserves as the authorities are forced to keep the stock of money unchanged at its desired level to maintain the exchange rate parity.

The authorities are forced to maintain a stock of reserves that will provide domestic residents with their desired money holdings, given the domestic source component. This implies that the commercial banks and the public will have their desired stock of base money. It also means that the authorities can effectively control the stock of official foreign exchange reserves by manipulating the domestic source component---such changes in the stock of reserves can be brought about at no cost in terms of price level changes, or output and employment changes when the price level is fixed. As you should have learned in the previous two lessons, output, income and prices are determined by the conditions of flow equilibrium.

It is evident from Equation 3 above that we must distinguish between two types of balance of payments disequilibria---stock and flow. A one-shot adjustment of the domestic source component or shift in the demand for base money holdings at a moment in time will lead to a shift in the stock of official reserves at that moment in time. The stock of official reserves will typically also be growing or declining at some rate through time---it is this flow of increases or decreases in the stock of reserves that is commonly referred to as the balance of payments surplus or deficit, which can be expressed as

4. $\Delta R = \Delta[(1/mm) \ P[\gamma - \theta(r* + \tau) + \varepsilon Y]] - \Delta Dsc$

where Δ is the change per unit time. Official reserve holdings change through time because the levels of income, prices, and the real interest rate and the money multiplier change through time---the effects of these changes are captured by the major collection of terms in the square brackets to the right of the equal sign. They also change as a result of a change in the domestic source component ΔDsc through time.

Notice that the presence of balance of payments equilibrium is completely independent of the condition of flow equilibrium---indeed, we did not have to include the IS equation in the above discussion. This is the case as long as people are free to buy and sell assets in the international market. There is no relationship of balance of payments equilibrium to equilibrium in the balance of trade---balance of payments equilibrium is entirely a monetary phenomenon. All this changes when we assume that there is zero private international mobility---that private residents are prohibited from purchasing assets from or selling them to foreign residents. In this case, the only domestically held foreign assets are official foreign exchange reserves and the condition of balance of payments equilibrium becomes

$$\Delta R = B_T$$

where B_T is the balance of trade---the balance of payments surplus or deficit becomes equal to the balance of trade surplus or deficit. To maintain balance of payments equilibrium the government has to bring about changes in the balance of trade by bringing about changes in the price level or income and employment or allowing the nominal exchange rate to change.

When there is international private capital mobility, the government can control the time-path of foreign exchange reserves simply by controlling the time-path of Dsc. But shocks to the demand for money are unpredictable and the main adjustment to these shocks will necessarily be the day-to-day purchases and sales of foreign exchange reserves in return for domestic currency necessary to keep the exchange rate at its fixed parity. Changes in the domestic source component perform the role of providing for long-run growth in the money supply to match the growth in demand as income and the volume of transactions rise with time.

Without growth in the domestic source component, the stock of foreign exchange reserves would grow without limit as the country's income and demand for money grows. This growth of reserves must be controlled because short-term foreign government securities, the main assets held as official reserves, are not a particularly good form for a country to hold large quantities of its wealth. Given the country's employed capital stock and the amount of wealth its residents possess, the bigger the fraction of that wealth held in the form of short-term foreign assets, the smaller will be the fraction held in equity and fixed income claims against domestically employed capital which will yield much higher returns than the treasury bills that form the greater part of the stock of official foreign exchange reserves.

The Canadian period of fixed exchange rates between late-1962 and mid-1970 provides an interesting example of the growth of official reserves in relation to base money. Figure 12.1 presents the two series expressed in millions of Canadian dollars.
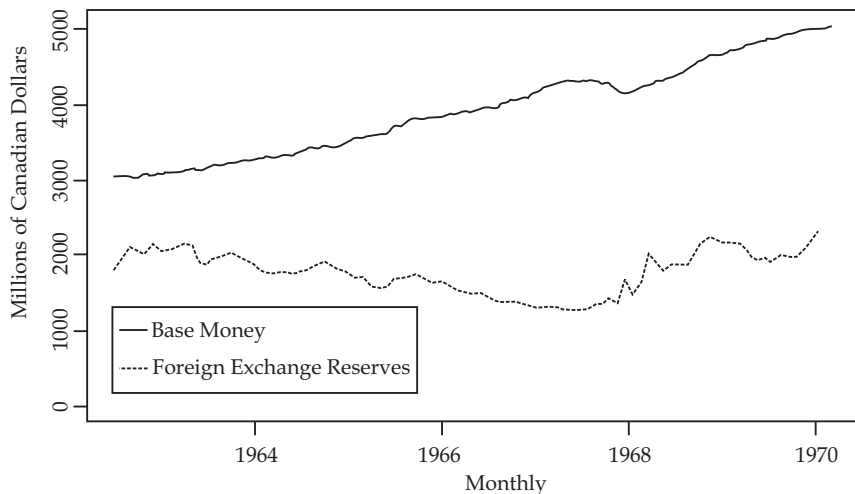
Figure 12.1 : Canada : Base Money and Foreign Exchange Reserves

Base money grew steadily in Canada except for a slight dip in late-1967 and early 1968. The Canadian stock of official foreign exchange reserves, on the other hand, declined very gradually until late-1967 and then trended upward thereafter, increasing sharply in 1970. The month-to-month changes in Canadian base money and official foreign exchange reserves are plotted in Figure 12.2.
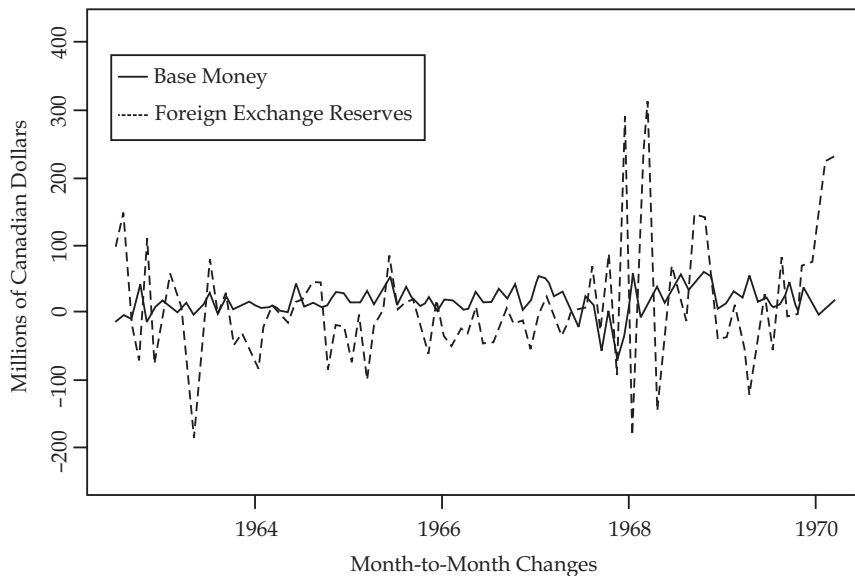
Figure 12.2 : Canada : Month-to Month Changes in Base Money and Official Reserves

The month-to-month changes in official foreign exchange reserves were much more variable than the month to month changes in base money, particularly in 1968, and reserves grew very sharply in early 1970. How do we explain this ?

The fact that reserves were more variable than base money suggests that these changes in official reserves were driven in considerable part by changes in the opposite direction in the domestic source component. Since H equals R plus Dsc , greater variability of R than H can only result from variability of Dsc in the opposite directions. An ordinary least squares regression of the month-to-month changes in foreign exchange reserves on the month-to-month changes in base money indicates no statistically significant relationship between them. The slope coefficient is negative with a P-Value of .23 --- indicating a 23 percent chance of observing a negative value of the magnitude observed purely on the basis of random chance when the true value is in fact zero---and the R-Square is only .01.

We have to conclude that much of the variability of the stock of official reserves was the result of the Bank of Canada's manipulation of the domestic source component but we should not venture a conclusion as to why the Bank was doing this without much more careful study. It is well-known that the Canadian Government abandoned the fixed exchange rate in mid-1970 in order to allow the government to control increasing upward domestic inflationary pressure which it was powerless to control under a fixed exchange rate. Accordingly, the observed escalating increases in the stock of official reserves in 1970 may well be the result of a fruitless attempt by the Bank of Canada to get a handle on domestic inflation by reducing the growth of the domestic source component of high-powered money. The only way to get control was to let the Canadian dollar float freely in the international market---only then could monetary policy become effective.

With respect to Canada's abandonment of the fixed exchange rate, it is useful to look at the movements in her real exchange rate with respect to the U.S., plotted in Figure 12.3.
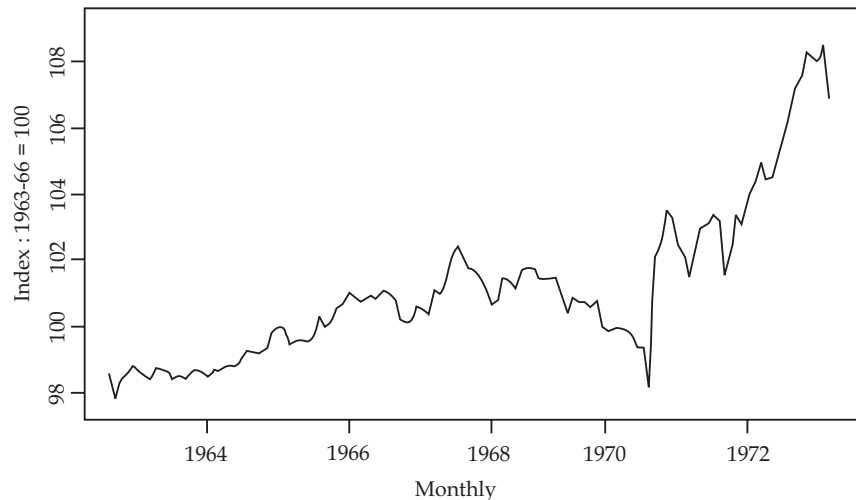
Figure 12.3 : Canadian Real Exchange Rate With Respect to United States

Recall from the definition of the real exchange rate that

$$P = Q \Pi P *.$$

If the government fixes the nominal exchange rate, the Canadian price level will vary up and down relative to the price level in the U.S. in response to movements in the equilibrium level of the real exchange rate Q arising from shifts in desired exports relative to imports, shifts in domestic relative to U.S. consumption and investment, changes in commodity prices, and so forth. When the exchange rate is fixed the Bank of Canada can not use monetary policy to avoid these price level changes.

Notice from Figure 12.3 that the real exchange rate increased by about 3.5 percent between mid-1962 and mid-1967--- an average rate of of increase of a bit less than 3/4 of a percent per year. Then between mid-1967 and mid-1970 the real exchange rate fell by about the same amount---at an average

rate of about 1 percent per year. One would not be surprised if the Bank of Canada tried to offset this decline, particularly the sharp decline in late-1967 and early 1968, by manipulation of the domestic source component. In this respect, you should keep in mind that the basic theoretical framework that we are applying here was not understood at that time, apart from the special case where there was perfect capital mobility, then interpreted as an extreme situation where domestic and foreign assets are identical. The instability of the month-to-month changes in the stock of official foreign exchange reserves during 1968 may have reflected failed attempts by the Bank of Canada to control the stock of base money. Finally, notice the 5 percent increase in the real exchange rate that occurred over June, July and August of 1970. Were the nominal exchange rate held fixed, this would have implied a 5 percent increase in Canada's equilibrium price level over the three month period. It is not surprising that the country abandoned the fixed exchange rate. Indeed, the real exchange rate increased by an additional 5 to 6 percent between late-1970 and late-1972, making the abandonment of the fixed exchange rate a wise decision.

Despite these Canadian hassles, the above discussion of balance of payments disequilibrium seems tame in comparison to what one reads in the popular press. One reads about international currency crises as edgy investors shift funds from currency to currency, creating havoc with the payments system. In fact, most of the time things are very quiet. But occasionally the ability of a country to maintain its fixed exchange rate parity comes into question. Major movements in a country's real exchange rate, as demonstrated for Canada, particularly downward ones that by pressing down on the domestic price level will lead to increased unemployment, are a good reason to expect abandonment of a fixed exchange rate. Another very real possibility is that the government is under political pressure to finance some of its expenditures by printing money---that is, having the central bank buy bonds from the treasury which will use the money so obtained for public expenditure. Or it may be that the government is under pressure to expand the money supply to deal with a difficult domestic unemployment situation in the face of a forthcoming election. Whatever its cause, such monetary expansion is inconsistent with maintaining the nominal exchange rate fixed. It becomes a reasonable bet that the country's currency will devalue in the future.

The spectre of a potential future devaluation presents big expected profit opportunities. If the domestic currency in fact devalues there are enormous gains to having one's assets denominated in foreign currencies while if devaluation does not occur little is lost by holding these currencies. Not surprisingly, huge shifts of funds out of domestic currency denominated and into foreign currency denominated assets occur under these conditions. To maintain the exchange rate in the face of these speculative pressures, the authorities must sell large quantities of foreign exchange reserves. This seems to put them in danger of eventually running out of reserves, in which case devaluation becomes inevitable. As official reserve holdings continue to fall, the pressure mounts. Of course, the country's authorities can easily, and costlessly in terms of effects on employment and prices, avoid running out of reserves---all they have to do is reduce the domestic source component of the money supply, something they may be reluctant to do under less-than-full-employment conditions.

Although governments can borrow official reserves from the governments of other countries in emergency circumstances or, better still, costlessly create them by reducing the domestic source component of base money, they must eventually put their house in order. This means either abandoning the fixed exchange rate, thereby proving the speculators to be right, or establishing credibility in the eyes of asset holders that the domestic money supply will be allowed to be determined endogenously at the fixed exchange rate, regardless of the real exchange rate shocks and public finance demands that may arise.

Speculative pressures on the exchange rate sometimes also arise under flexible exchange rates. When investors think that a currency will depreciate in the future, they will shift funds out of assets denominated in that currency now. This means that the prices of those assets will fall and interest rates on them will rise to reflect a forward discount on the currency expected to depreciate. Under these circumstances, central banks may "lean against" movements in the external value of their currencies by purchasing and selling foreign exchange. Balance of payments disequilibria can thus arise even when the exchange rate is flexible although these are of minor consequence and probably should not even be referred to as disequilibria.

An additional issue should be mentioned before we proceed to the test. It should be obvious that sterilization of the effects of changes in official holdings of foreign exchange reserves on the money supply is impossible under fixed exchange rates when assets can be freely bought and sold across international boundaries. Suppose that the demand for nominal money holdings declines and R falls. Any attempt of the authorities to offset this fall in R by an increase in Dsc will lead to a further fall in R equal to that increase in Dsc. The government has no control over the domestic money supply under fixed exchange rates.

It is time for a test. Figure out your own answers to the questions before looking at the ones provided.

Did u know? The difference between speculative pressures under flexible and fixed exchange rates is that under flexible exchange rates it is harder to guess which way the rate is likely to move in the future. When speculative pressures arise under fixed exchange rates it is usually quite clear in which direction, if any, the exchange rate will move.

Types of BOP Equilibrium

There are two types of BOP equilibrium, i.e., static equilibrium and dynamic equilibrium :

(a) **Static Equilibrium :** The distinction between static and dynamic equilibrium depends upon the time period. In static equilibrium, exports equal imports including exports and imports of services as well as goods and the other items on the BOPs – short term capital, long term capital and monetary gold are on balance, zero. Not only should the BOPs be in equilibrium, but also national money incomes should be in equilibrium vis-a-vis money incomes abroad. The foreign exchange rate must also be in equilibrium.

(b) **Dynamic Equilibrium :** The condition of dynamic equilibrium for short periods of time is that exports and imports differ by the amount of short-term capital movements and gold (net) and there are no large destabilising short-term capital movements.

The **condition** for dynamic equilibrium in the long run is that exports and imports differ by the amount of long term autonomous capital movements made in a normal direction, i.e. from the low-interest rate country to those with high rates. When the BOP of a country is in equilibrium, the demand for domestic currency is equal to its supply. The demand and supply situation is thus neither favourable nor unfavourable. If the BOP moves against a country, adjustments must be made by encouraging exports of goods, services or other forms of exports or by discouraging imports of all kinds. No country can have a permanently unfavourable BOP, though it is possible – and is quite common for some countries – to have a permanently unfavourable balance of trade. Total liabilities and total assets of nations, as of individuals, must balance in the long-run.

12.2 Disequilibrium in Balance of Payments

Meaning of Disequilibrium in Balance of Payment

Though the credit and debit are written balanced in the balance of payment account, it may not remain balanced always. Very often, debit exceeds credit or the credit exceeds debit causing an imbalance in the balance of payment account. Such an imbalance is called the disequilibrium. Disequilibrium may take place either in the form of deficit or in the form of surplus.

Disequilibrium of **Deficit** arises when our receipts from the foreigners fall below our payment to foreigners. It arises when the effective demand for foreign exchange of the country exceeds its supply at a given rate of exchange. This is called an 'unfavourable balance'. Disequilibrium of **Surplus** arises when the receipts of the country exceed its payments. Such a situation arises when the effective demand for foreign exchange is less than its supply. Such a surplus disequilibrium is termed as 'favourable balance'.

Causes of Disequilibrium in Balance of Payment

1. **Population Growth**

   Most countries experience an increase in the population and in some like **India** and **China** the population is not only large but increases at a faster rate. To meet their needs, imports become essential and the quantity of imports may increase as population increases.

2. **Development Programmes**

   Developing countries which have embarked upon planned development programmes require to import capital goods, some raw materials which are not available at home and highly skilled and specialized manpower. Since development is a continuous process, imports of these items continue for the long time landing these countries in a balance of payment deficit.

3. **Demonstration Effect**

   When the people in the less developed countries imitate the consumption pattern of the people in the developed countries, their import will increase. Their export may remain constant or decline causing disequilibrium in the balance of payments.

4. **Natural Factors**

   Natural calamities such as the failure of rains or the coming floods may easily cause disequilibrium in the balance of payments by adversely affecting agriculture and industrial production in the country. The exports may decline while the imports may go up causing a discrepancy in the country's balance of payments.

5. **Cyclical Fluctuations**

   Business fluctuations introduced by the operations of the **trade** cycles may also cause disequilibrium in the country's balance of payments. For example, if there occurs a business recession in foreign countries, it may easily cause a fall in the exports and exchange earning of the country concerned, resulting in a disequilibrium in the balance of payments.

6. **Inflation**

   An increase in income and price level owing to rapid **economic development** in developing countries, will increase imports and reduce exports causing a deficit in balance of payments.

7. **Poor Marketing Strategies**

   The superior marketing of the developed countries have increased their surplus. The poor marketing facilities of the developing countries have pushed them into huge deficits.

8. **Flight of Capital**

   Due to speculative reasons, countries may lose foreign exchange or gold stocks People in developing countries may also shift their capital to developed countries to safeguard against political uncertainties. These capital movements adversely affect the balance of payments position.

9. **Globalisation**

   Due to globalisation there has been more liberal and open atmosphere for international movement of goods, services and capital. Competition has beer increased due to the globalisation of international economic relations. The emerging new global economic order has brought in certain problems for some countries which have resulted in the balance of payments disequilibrium.

Types of BOP Disequilibrium

There are three main types of BOP Disequilibrium which are discussed below :

(a)   Cyclical disequilibrium,

(b)   Secular disequilibrium, and

(c)   Structural Disequilibrium.

(a) **Cyclical Disequilibrium :** Cyclical disequilibrium occurs because of **two** reasons. *First two countries may be passing through different paths of business cycle. Second, the countries may be following the same path but the income elasticities of demand or price elasticities of demand are different.* If prices rise in prosperity and decline in depression, a country with a price elasticity for imports greater than unity will experience a tendency for decline in the value of imports in prosperity; while those for which import price elasticity is less than one will experience a tendency for increase. (These tendencies may be overshadowed by the effects of income changes, of course. Conversely, as prices decline in depression, the elastic demand will bring about an increase in imports, the inelastic demand a decrease.)

(b) **Secular Disequilibrium :** *The secular or long-run disequilibrium in BOP occur because of long-run and deep seated changes is an economy as it advances from one stage of growth to another.* (The current account follows a varying pattern from one state to another. In the initial stages of development, domestic investment exceeds domestic savings and imports exceed exports.

Disequilibrium arises owing to lack of sufficient funds available to finance the import surplus, or the import surplus is not covered by available capital from abroad. Then comes a stage when domestic savings tend to exceed domestic investment and exports outrun imports. Disequilibrium may result, because the long-term capital outflow falls short of the surplus savings or because surplus savings exceed the amount of investment opportunities abroad. At a still later stage, domestic savings tend to equal domestic investment and long term capital movements are on balance, zero.)

(c) **Structural Disequilibrium :** Structural disequilibrium can be further bifurcated into :

   (i) **Structural Disequilibrium at Goods Level :** *Structural disequilibrium at goods level occurs when a change in demand or supply of exports or imports alters a previously existing equilibrium, or when a change occurs in the basic circumstances under which income is earned or spent abroad, in both cases without the requisite parallel changes elsewhere in the economy.* (Suppose the demand for Pakistani handicrafts falls off. The resources engaged in the production of these handicrafts must shift to some other line or the country must restrict imports, otherwise the country will experience a structural disequilibrium.

   A deficit arising from a structural change can be filled by increased production or decreased expenditure, which in turn affect international transactions in increased exports or decreased imports. Actually it is not so easy, because the resources are relatively immobile and expenditure not readily compressible. Disinflation or depreciation may be called for to correct a serious disequilibrium.)

   (ii) **Structural Disequilibrium at Factors Level :** *Structural disequilibrium at the factor level results from factor prices which fall to reflect accurately factor endowments, i.e., when factor prices are out of line with factor endowments, distort the structure of production from the allocation of resources which appropriate factor prices would have indicated.* If, for instance, the price of labour is too high, it will be used more sparingly and the country will import goods with a higher labour content. This will lead to unemployment, upsetting the balance in the economy.

Self-Assessment

1. **Choose the correct options:**

   (*i*) Which of the following would be one of the results associated with the use of freely floating foreign exchange rates to correct a nation's balance of payments surplus?

   (*a*) The nation's terms of trade with other nations would be worsened.

   (*b*) Importers in the nation who had made contracts for the future delivery of goods would find that they had to pay a higher price than expected for the goods.

   (*c*) If the nation were at full employment, the decrease in exports and the increase in imports would be inflationary.

   (*d*) Exporters in the nation would find their sales abroad had decreased.

LOVELY PROFESSIONAL UNIVERSITY

(*ii*) When exchange rates are fixed and a nation at full employment has a balance of payments surplus, the result in that nation will be

    (*a*) a declining price level   (*b*) falling currency income

    (*c*) inflation   (*d*) rising real income

(*iii*) The use of exchange controls to eliminate a nation's balance of payments deficit results in decreasing the nation's

    (*a*) imports   (*b*) exports

    (*c*) price level   (*d*) income

(*iv*) Which condition did a nation have to fulfill if it were to be under the gold standard?

    (*a*) use only gold as a medium of exchange

    (*b*) maintain a flexible relationship between its gold stock and its currency supply

    (*c*) allow gold to be freely exported from and imported into the nation

    (*d*) define its monetary unit in terms of a fixed quantity of dollars

(*v*) If the nations of the world were on the gold standard and one nation has a balance of payments surplus,

    (*a*) foreign exchange rates in that nation would rise

    (*b*) gold would tend to be imported into that country

    (*c*) the level of prices in that country would fall

    (*d*) employment and output in that country would fall

(*vi*) Which was the principal disadvantage of the gold standard?

    (*a*) unstable foreign exchange rates

    (*b*) persistent payments imbalances

    (*c*) the uncertainties and decreased trade that resulted from the depreciation of gold

    (*d*) the domestic macroeconomic adjustments experienced by a nation with a payments deficit or surplus

12.3 Summary

- The main purpose of this unit is to arrive at definitions of balance of payments equilibrium sufficiently precise to furnish policy-makers with meaningful guides.

- There is no doubt that a study of country's balance of payment reveals much information about its economic position and development of the country. But when we are to see that a country is heading towards financial bankruptcy or higher standard of living, we are to examine the balance of payments of many years of that country.

- Balance of payments equilibrium occurs when induced balance of payments transactions--- those engineered by the government to influence the nominal exchange rate---are zero. This implies that autonomous receipts from exports and the sale of securities abroad equal autonomous payments for imports and the purchase of securities from foreign residents. Since changes in the stock of official reserves of foreign exchange are the method used by the authorities to fix or otherwise manipulate the exchange rate, balance of payments equilibrium requires that the stock of foreign exchange reserves be constant.

- **Static Equilibrium :** The distinction between static and dynamic equilibrium depends upon the time period. In static equilibrium, exports equal imports including exports and imports of services as well as goods and the other items on the BOPs – short term capital, long term capital and monetary gold are on balance, zero.

- **Dynamic Equilibrium :** The condition of dynamic equilibrium for short periods of time is that exports and imports differ by the amount of short-term capital movements and gold (net) and there are no large destabilising short-term capital movements.

- Though the credit and debit are written balanced in the balance of payment account, it may not remain balanced always. Very often, debit exceeds credit or the credit exceeds debit causing an imbalance in the balance of payment account. Such an imbalance is called the disequilibrium. Disequilibrium may take place either in the form of deficit or in the form of surplus.

12.4 Key-Words

1. Equilibrium	:	In economics, economic equilibrium is a state of the world where economic forces are balanced and in the absence of external influences the (equilibrium) values of economic variables will not change. For example, in the standard text-book model of perfect competition, equilibrium occurs at the point at which quantity demanded and quantity supplied are equal.
2. Structural Disequilibrium	:	A situation where internal and/or external forces prevent market equilibrium from being reached or cause the market to fall out of balance. This can be a short-term byproduct of a change in variable factors or a result of long-term structural imbalances.

12.5 Review Questions

1. Write a short note on the equilibrium of balance of payments.
2. What are the types of balance of payment equilibrium ? Discuss.

Answers: Self-Assessment

1. (*i*) (*d*)   (*ii*) (*c*)   (*iii*) (*a*)   (*iv*) (*c*)   (*v*) (*b*)
    (*vi*) (*d*)

12.6 Further Readings

Books

1. Adapted, with slight modifications, from the definition given in the International Monetary Fund *Balance of Payments Manual*, 2nd edn., January 1950, *p.* 1.
2. Leland B. Yeager, *International Monetary Relations* (New York : Harper & Row, 1966), *p.* 36.
3. J.E. Meade, *The Theory of International Economic Policy*, *Vol. I : The Balance of Payments*, (1951) *p.* 7. See also Yeager, *op. cit.*, *p.* 452.
4. W.M. Scammel, *International Trade and Payments* (Toronto : The Macmillan Company, 1974) *p.* 272.
5. J.E. Meade, *op. cit.*, *p.* 11.

Unit 13: BOP Adjustment : Monetary Approach

```
CONTENTS

Objectives

Introduction

13.1  Monetary Approach to BOP Adjustments

13.2  Elasticity Approach

13.3  Summary

13.4  Key-Words

13.5  Review Questions

13.6  Further Readings
```

Objectives

After reading this Unit students will be able to:

•   Understand BOP Adjustment.

•   Explain the Monetary Approach to BOP Adjustment.

Introduction

The BOP adjustment has been a theoretically complex and practically complicated issue for both the economists and the policy maker. The economists have suggested variety of approaches and measures to correct the BOP disequilibrium. Correcting disequilibrium in the BOP has been, as noted above, a knotty problem for both economists and policy-makers. However, efforts continued to find a general solution to the problem of BOP adjustment. In the process, there emerged another important approach to BOP adjustment, called *monetary approach.* It must be borne in mind that the monetary approach developed by these economists is different from monetary-policy approach. The monetary approach to BOP adjustments is discussed below.

13.1 Monetary Approach to BOP Adjustments

According to the modern monetary approach. BOP, *disequilibrium is a monetary phenomenon. The BOP disequilibrium (deficit or surplus) arises because of inbalance between the demand for and supply of money.* The BOP **deficits** arise because money supply exceeds the demand for money and BOP **surplus** is the outcome of the excess of demand for money over the supply of money. The monetary approach is based on the premise that BOP disequilibrium arising out of discrepancy between the demand for and supply of money is a transitory phenomenon and is self-correcting in the long run.

Another important feature of the monetary approach is that it looks at the BOP "bottoms up." That is, it begins the analysis of BOP disequilibrium by looking at the bottom of the BOP accounts, *i.e.,* the change in the official foreign exchange reserves. Rather than analyzing 'above the line' entries in the BOP accounts, monetary approach lumps together all the transactions in the current and capital accounts and looks at the changes in the final outcome of the international transactions. The final outcome is reflected by the change in the official reserves of foreign exchange.

Pending the details, let us first look at the process of *self-correction.* The self-correction process begins with the change in the *official reserves.* In case there is deficit in the BOP, it causes a decline in the official reserves of foreign exchange. Decline in the foreign exchange reserves leads to a decline in the money supply. The decrease in money supply leads to decrease in domestic prices, increase in exports, decrease in imports, and decrease in trade deficits. These trends automatically correct the BOP disequilibrium. On the other hand, a surplus in the BOP increases money supply which causes rise in

prices, decrease in exports, and rise in imports. This decreases the trade surplus. Thus, the surplus in BOP is automatically eliminated and BOP disequilibrium is corrected. These conclusions can be established through the basic model of the monetary approach.

> *Did u know?* The monetary approach was developed by Robert A. Mundell in 1968 and 1971 and Harry G. Johnson in 1972.

The Model of Monetary Approach

According to the monetary approach, BOP imbalance (B) equals the difference between the demand for money (M_d) and the supply of money (M_s).

$$B = M_d - M_s \qquad \qquad ... (1)$$

The relationship between the BOP and money demand and supply is illustrated in Figure 1. In this figure, the demand for money, M_d, is assumed to be exogenously determined and to remain constant. This assumption can be explained as follows. We know that,

$$M_d = M_t + M_{sp} \qquad \qquad ... (2)$$

where, M_t = transaction demand, and M_{sp} = speculative demand for money.

We know also that $M_t = f(Y)$ and $M_{sp} = f(i)$. where Y = national income, and i = interest rate. The monetary approach to BOP adjustment assumes that both Y and i are exogenously determined. It implies that so long as Y and i remain constant, the demand for money remains constant as shown by the schedule M_d in Figure 1.

As regards the supply of money, monetary approach assumes that money supply in an open economy equals *domestic component of money supply* plus *external component of money supply.* The domestic component of money supply equals *money multiplier* times the commercial bank reserves with the central bank (*i.e.*, the domestic component of the monetary base). The external component of money supply equals money multiplier times the 'international reserves' (*i.e.*, the external component of the monetary base). Given these components of money supply, M_s, in Eq. l can be defined as :

$$M_s = m(DB + IB) \qquad \qquad ... (3)$$

where, m = money multiplier; DB = domestic base (commercial banks' deposits with the central bank); and IB = international base (international reserves).

The monetary approach to BOP adjustment assumes DB and money multiplier (m) to remain constant. Therefore, domestic component of money supply remains constant. The constant domestic component of money supply has been shown in Figure 13.1 by the horizontal schedule m(DB) by assuming mDB = Rs. 100 billion. The schedule m(DB) has been drawn by assuming $m = 5$ and DB = Rs. 20 billion (constant).
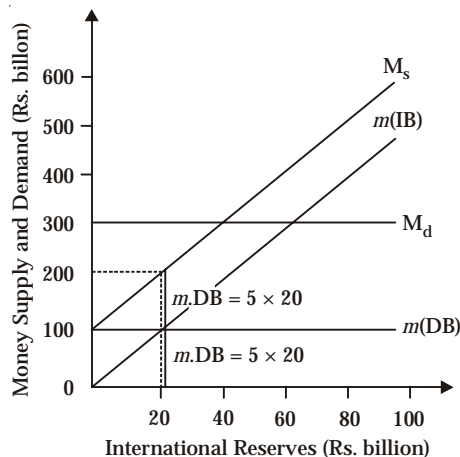
Figure 13.1 : Money Demand and Supply and International Balance

On the other hand, the international balance (IB) is supposed to be a variable factor. Therefore, the international component of money supply changes with the change in the international balance (IB). It equals $m \times$ IB. The international component of money supply schedule thus obtained is given by the schedule m(IB). The m(IB) schedules in our example is based on s(IB). The aggregate money supply schedule. M_s, is the vertical sum of schedules m(DB) and m(IB). This is precisely the *monetary approach* model. Let us now look at the self-correcting mechanism of the BOP disequilibrium as envisaged by the monetary approach to the problem.

The Self-Correcting Mechanism

The self-correcting mechanism under monetary approach is illustrated in Figure 13.2 assuming a fixed exchange rate. Panel (a) of this figure is reproduction of Figure 13. 1 and panel (b) is a derivation from panel (a). As shown in panel (a) the total money supply schedule (M_s) and the total money demand schedule (M_d) intersect at point E where M_d equals M_s at Rs. 300 billion and, at the equilibrium, international reserves equals Rs. 40 billion. It implies that if international reserves equal Rs. 40 billion, then M_d will always be equal to M_s and the monetary sector will be in equilibrium.

The point that needs to be noted here is that, according to the monetary approach, when total demand for money equals total money supply, then BOP is in equilibrium—there is neither surplus nor deficit in the BOP. As shown in panel (a) of Figure 13.2, at no point other than point E money supply equals money demand. Therefore, at all other points along the M_s and M_d schedules, there is either deficit or surplus in the BOP. This phenomenon is illustrated in panel (b) of Figure 13. 2. The *payments imbalance schedule* represents the vertical difference between the M_d and M_s schedules in panel (a). Point E' in panel (b) shows zero BOP balance corresponding to point E in panel (a).

The Self-Correcting Process : According to the monetary approach, the self-correcting process is an in-built system. When demand for money exceeds the supply of money, it means that there is surplus in the BOP. As Figure 13. 2 shows, at all the points below point E along the schedule M_s in panel (a), demand for money exceeds supply of money. It means that, in accordance with Eq. 1, there is BOP surplus, that is, increase in the foreign exchange reserves. For example, at foreign exchange reserves of Rs. 20 billion, the demand for money equals Rs. 300 billion and money supply equals Rs. 200 billion. It means that money demand exceeds money supply by JK = Rs. 100 billion. This means a BOP surplus of Rs. 100 billion, which enhances the foreign exchange reserves. This rise in the foreign exchange reserves pushes the money supply up over time along the schedule M_s towards point E. At point E, demand for money equals the supply of money. Therefore, there is neither surplus nor deficit in the BOP. That is, the BOP is automatically restored to equilibrium.
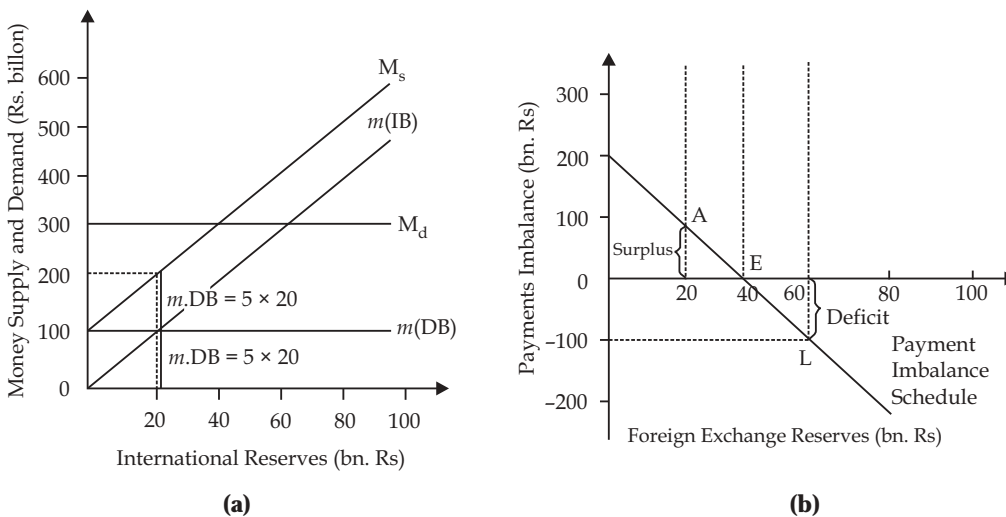
Figure 13.2 : Self-Correcting Monetary Mechanism

Similarly, when money supply exceeds money demand, BOP shows a deficit. For example, at foreign exchange reserves of Rs. 60 billion, money supply exceeds money demand by LM = Rs. 400 billion – Rs. 300 billion = Rs. 100 billion. This means a BOP deficits of Rs. 100 billion. The deficit in the BOP decreases the money supply over time. The decrease in money supply makes the system move back to point E. This process continues until the money supply falls to Rs. 400 billion to equal the money demand. At point E, the demand for money equals the supply of money. It means B = 0 [in Eq. (l)] and their is neither deficit nor surplus in the balance of payment. This is how the BOP equilibrium is automatically restored.

13.2 Elasticity Approach

The elasticity approach, which has been associated with Robinson (1937), places its emphasis on the effects of exchange rate changes on the exports and imports of a country and, hence, on the trade account balance, whilst ignoring all other variables such as income. The elasticity approach applies the Marshall-Lerner condition, which states that the sum of the elasticities of demand for imports and exports must be greater than unity in absolute terms for a devaluation to improve the balance of payments (Du Plessis et al., 1998).

The logic behind this condition is as follows. Suppose the elasticity of demand for exports is zero. In this case exports in domestic currency are the same as before devaluation. If the sum of the elasticities is greater than one, the elasticity of demand for imports must be greater than one, so that the value of imports falls. With no fall in the value of exports and a fall in the value of imports, the balance of payments improves. Now, suppose the demand for imports has zero elasticity. The value of imports will rise by the full percentage of devaluation. If the elasticity of demand for exports is greater than unity, the value of exports will expand by more than the percentage of devaluation. Therefore, the balance of payments will improve. If each element of the elasticity of demand is less than unity, but the sum is greater than unity, the balance of payments will improve because expansion of exports in domestic currency will exceed the value of imports.

Absorption Approach

The absorption approach was first presented by Alexander (1952). He sought to look at the balance of trade from the point of view of national income accounting. It is useful in pointing out that an improvement in the balance of trade calls for an increase in production relative to absorption. The absorption approach intends to show how devaluation might change the relationship between expenditures or between absorption and income - in both nominal and real terms. It is worth noting that great emphasis is laid on the current account balance. This approach contends that the devaluation of a currency would lead to an increase in inflationary prices, which would in turn revoke the initial effect of an increase in prices.

The starting point of the absorption approach is the national income identity: Y=C +I + G + X-M (5) Where Y = national income; C = private consumption of goods and services purchased at home and from abroad; I = total investment, by firms as well as by government; G = government expenditure on goods and services; X = exports of goods and services; and M = imports of goods and services. Combining C + I + G expenditure terms into a single term, A, representing domestic absorption (i.e., total domestic expenditure) and X - M terms into B, net exports/trade balance, we get: Y=A+B (6) equation 6 states that national income equals absorption plus the trade balance, or alternatively B=Y-A (7) Equation 7 states that the trade balance is equal to the difference between domestic income and total absorption. Equation (7) is the fundamental equation of the absorption approach. It implies that, if total absorption (expenditure) exceeds income (production), then imports will exceed exports, resulting in a balance of payments deficit. If the opposite occurs, i.e. where income exceeds absorption, then the balance of payments will be in surplus. A balance of payments deficit can, therefore, only be corrected if the level of absorption changes relative to the level of income (Du Plessis et al., 1998:251). The empirical literature is replete with studies on the monetary approach to balance of payment. Mixed results were obtained from the different studies on the MABP. Obioma (1998) used data for 1960-1993 to test the validity of monetary approach to the balanceof-payments adjustment for Nigeria

under fixed and flexible exchange regimes. He found that an increase in domestic credit or money stock leads to external reserves outflow or adverse balance of payments during the fixed exchange rate regime. But in the flexible exchange rate era, an increase in domestic credit brings about exchange rate depreciation. Using data for 1960-1995, Jimoh (1990) also found that the monetary approach is relevant in analyzing balance-of-payments adjustments in Nigeria.

The empirical results validate the MABP in Zimbabwe. The implication of such result is that money played a significant role in the determination of deficits in the balance of payments. In a related development, Coppin (1994:83) carried out a study for Barbados to test the validity of the MABP. It was evident from his results that the ?degree of openness of an economy? played a particularly important role in determining international reserves. He found strong evidence in support of the MABP in Barbados. Specifically, the result revealed that expansionary fiscal policy played a vital role over monetary factors in determining international reserves. Furthermore, Aghevli and Khan (1977) performed an empirical test on the MABP for 39 developing countries and found highly significant results, maintaining that the mechanisms underlying the MABP held strongly for these countries.

Lachman (1975) conducted a study to test the validity of the MABP for South Africa. His results found strong evidence in favour of the MABP, and concluded that monetary authorities would definitely be able to predict the extent to which increases in money supply would augment imports. Leon (1988), also examined the applicability of the MABP in Jamaican. He used the reserve-flow and sterilization equations using both single and simultaneous equations. The results found strong evidence in favour of the reserve-flow equation, and that the MABP's predictions were not rejected. The results also revealed that monetary authorities were in fact sterilising reserves in Jamaica. Watson (1990), also conducted a study on Trinidad and Tobago in assessing the MABP, using data for the period 1965-1985. The results did not find any support in favour of the MABP, indicating that balance of payment problems in Trinidad and Tobago is not a monetary policy issue.

In summary, there is convincing evidence that the MABP in fact is an important concept In the literature and an unresolved issue. While some studies found evidence of the MABP (see Lachman (1975), Dhliwayo (1996)), others including Watson (1990), did not found any support in favour of the MABP. Most parts of the empirical literature were based on the ?reserve-flow equation', where a country's international reserves, or the rates of change in reserves, are regarded as the dependent variable. On the other hand, the independent variables vary in the different studies. They can include domestic income, prices, the interest rate, government expenditure, money multiplier, money stock, the exchange rate, and demand for nominal and real money balances. Thus, given the conflicting evidence in the literature and the prevalence of the BOP deficit in the WAMZ countries, it is necessary to undertake an empirical study in order to validate the MABP.

Notes Dhliwayo (1996) used data for the period 1980-1991 to investigate the MABP in Zimbabwe. His findings indicate a one-to-one negative relationship between domestic credit and the flow of international reserves.

Self-Assessment

1. **Choose the correct options:**

    (*i*) Balance of Payments equilibrium means

       (*a*) the balance of trade equals zero

       (*b*) the current account balance equals zero

       (*c*) the capital account balance equals zero

       (*d*) the current account balance is exactly offset by the capital account balance

Notes

(*ii*) Macroeconomic equilibrium in an open economy requires

    (*a*) goods market equilibrium

    (*b*) money market equilibrium

    (*c*) balance of payments equilibrium

    (*d*) all of the above

    (*e*) only (*a*) and (*b*) above

(*iii*) An implication of the asset approach is that

    (*a*) since financial asset prices change rapidly, exchange rate should vary more than goods prices

    (*b*) since PPP holds even in the short-run, goods prices should change as rapidly as the exchange rate

    (*c*) covered interest parity will rarely hold if there is perfect capital mobility

    (*d*) goods prices will be largely unaffected by exchange rate changes in the long run

(*iv*) Suppose the U.S. fixes the price of gold at \$2.00 per ounce, and Germany fixes the price of gold at €4.00 per ounce. The \$/€ exchange rate is

    (*a*) 1:2   (*b*) 2:1

    (*c*) 1:1   (*d*) 1:4

(*v*) The elasticity approach to the balance of trade

    (*a*) focuses on the effects of changing relative prices of domestic and foreign goods on the balance of trade

    (*b*) is only applicable with a floating exchange rate system

    (*c*) is based on an analysis of the absolute prices of a country's exports in world markets

    (*d*) indicates that the demand elasticity for exports is always inelastic

(*vi*) The J-curve

    (*a*) only occurs under fixed exchange rates

    (*b*) occurs when a currency is appreciated

    (*c*) shows the initial worsening of the balance of trade with a devaluation

    (*d*) shows how the balance of trade immediately improves after a devaluation

(*vii*) The monetary approach to the balance of payments

    (*a*) focuses on the importance of the current account balance

    (*b*) views exchange rate variations as the sole adjustment mechanism

    (*c*) highlights the role of financial assets of payments adjustment

    (*d*) implies that the price level and exchange rate can both be determined by the central bank

13.3 Summary

- The main aim of this paper was to determine the dynamics of balance of payments adjustment in the countries of the West African Monetary Zone (WAMZ). Specifically, the study seeks to examine the MABP theory and its implication for the WAMZ countries. This entails estimating a net foreign asset equation and testing if the estimated partial coefficient of changes in domestic credit with respect to net foreign asset is not significantly different from minus one. If the condition is satisfied, the inference is that the money plays a role in the determination of balance of payments deficits. The empirical results confirm that money has played a significant role in the determination of deficits in the balance of payments. Thus, the findings from both the within-country and cross-country effects suggest that the monetary approach to the WAMZ countries balance of payments is indeed applicable. Specifically, the findings indicate that both interest

rate and domestic credit had negative relationship with net foreign asset with statistically significant coefficients. In addition, the result established a positive relationship between log of GDP and net foreign asset with statistically significant coefficients.

- A major policy implication of this study is that the monetary approach to the balance of payments holds in the WAMZ countries studied, since growth in domestic credit is an important determinant of their balance of payments position. Therefore, a tight rein on domestic credit creation is a necessary condition for maintaining stability in the balance of payments over time. Thus, monetary authorities should pay special attention to domestic credit creation when controlling the country's balance of payments. It is important that the country achieves sufficient economic growth through money demand to correct the balance of payments deficit. Authorities of the WAMZ countries should also look at the increased budget deficit, which is mostly financed through borrowing from the central bank. The expansion in the fiscal deficit caused the increases in domestic credit.

- Another policy implication is the need to manage domestic liquidity wisely in view of the tremendous pressure on the balance of payments of excess money. A determined effort to mobilize resources through private saving and the implementation of a prudent fiscal policy through efficient collection of tax revenues, rationalization of government expenditure towards growth enhancing and poverty reduction programmes will also enable the government to pursue its development programs without having to rely on the monetization of its budget deficit.

13.4 Key-Words

1. Monetary Approach : A framework for analyzing exchange rates and the balance of payments that focuses on supply and demand for money in different countries.

2. Deficit : A deficit is the amount by which a sum falls short of some reference amount. In economics, a deficit is a shortfall in revenue

13.5 Review Questions

1. What is the monetary approach to BOP Adjustments?

2. Discuss the model of Monetary approach.

3. Distinguish between elasticity and absorption approach.

Answers: Self-Assessment

1. (*i*) (*d*)   (*ii*) (*d*)   (*iii*) (*a*)   (*iv*) (*a*)   (*v*) (*a*)
   (*vi*) (*c*)   (*vii*) (*c*)

13.6 Further Readings

Books

1. Alexander, S. S., "Effects of Devaluation on the Trade Balance," *IMF Staff Papers,* Vol.2, April 1952, reprinted in H.G. Johnson and R. E. Caves (eds). *Readings in International Economics,* (Richard D. Irwin, Homewood, III., 1968).

2. Chacholiades, M., *International Economics,* (McGraw-Hill, New York, 1990), Chs. 12-15 and 19.

3. Demburg, Thomas, F., *Macroeconomics : Concepts, Theories and Policies,* (McGraw-Hill, New York, 1985). Chs. 15 and 16.

4. Dombusch, D, Fischer, S., and Startz, R., *Macroeconomics,* (McGraw-Hill, New York, 7th edn.), Ch. 12. International Monetary Fund (IMF), *Balance of Payments Mannual,* 4th Edn.(1977).

Unit 14 : Exchange Rate : Meaning and Components

CONTENTS

Objectives

Introduction

14.1 Meaning of Exchange Rate

14.2 Components of Exchange Rate

14.3 Summary

14.4 Key-Words

14.5 Review Questions

14.6 Further Readings

Objectives

After reading this Unit students will be able to:

•   Know the Meaning of Exchange Rate.

•   Discuss the Components of Exchange Rate.

Introduction

In virtually all modern economies, money (i.e. currency) is created and controlled by a central governing authority. In most cases, currencies are developed by individual countries, though this need not be the case. (One notable exception is the Euro, which is the official currency for most of Europe.) Because countries buy goods and services from other countries (and sell goods and service to other countries), it's important to think about how currencies of one country can be exchanged for currencies of other countries.

Like other markets, foreign-exchange markets are governed by the forces of supply and demand. In such markets, the "price" of a unit of currency is the amount of another currency that is needed to purchase it. For example, the price of one Euro is, as of the time of writing, about 1.25 US dollars, since currency markets will exchange one Euro for 1.25 US dollars.

These currency prices are referred to as exchange rates. More specifically, these prices are *nominal exchange rates* (not to be confused with real exchange rates). Just as the price of a good or service can be given in dollars, in Euro, or in any other currency, an exchange rate for a currency can be stated relative to any other currency. You can see a variety of such exchange rates by going to various finance web sites.

A US Dollar/Euro (USD/EUR) exchange rate, for example, gives the number of US dollars than can be bought with one Euro, or the number of US dollars per Euro. In this way, exchange rates have a numerator and a denominator, and the exchange rate represents how much numerator currency can be exchanged for one unit of denominator currency.

Because currency prices are stated relative to another currency, economists say that currencies appreciate and depreciate specifically relative to other currencies.

Appreciation and depreciation can be inferred directly from exchange rates. For example, If the USD/EUR exchange rate were to go from 1.25 to 1.5, the Euro would buy more US dollars than it did before. Therefore, the Euro would appreciate relative to the US dollar. In general, if an exchange rate increases, the currency in the denominator (bottom) of the exchange rate appreciates relative to the currency in the numerator (top).

> Changes in the price of a currency are referred to as appreciation and depreciation. Appreciation occurs when a currency becomes more valuable (*i.e.* more expensive), and depreciation occurs when a currency becomes less valuable (*i.e.* less expensive).

Similarly, if an exchange rate decreases, the currency in the denominator of the exchange rate depreciates relative to the currency in the numerator. This concept can be a little tricky, since it's easy to get backwards, but it makes sense : for example, if the USD/EUR exchange rate were to go from 2 to 1.5, a Euro buys 1.5 US dollars rather than 2 US dollars. The Euro therefore depreciates relative to the US dollar, since a Euro doesn't trade for as many US dollars as it used to.

14.1 Meaning of Exchange Rate

The foreign exchange rate or exchange rate is the rate at which one currency is exchanged for another. It is the price of one currency in terms of another currency. It is customary to define the exchange rate as the price of one unit of the foreign currency in terms of the domestic currency. The exchange rate between the dollar and the pound refers to the number of dollars required to purchase a pound. Thus the exchange rate between the dollar and the pound from the US viewpoint is expressed as $ 2.50 = £ 1. The Britishers would express it as the number of pounds required to get one dollar, and the above exchange rate would be shown as £ 0.40 = $ 1.

The exchange rate of $ 2.50 = £ 1 or £ 0.40 = $ 1 will be maintained in the world foreign exchange market by arbitrage. *Arbitrage refers to the purchase of a foreign currency in a market where its price is low and to sell it in some other market where its price is high.* The effect of arbitrage is to remove differences in the foreign exchange rate of currencies so that there is a single exchange rate in the world foreign exchange market. If the exchange rate is $ 2.48 in the London exchange market and $ 2.50 in the New York exchange market, foreign exchange speculators, known as arbitrageurs, will buy pounds in London and sell them in New York, thereby making a profit of 2 cents on each pound. As a result, the price of pounds in terms of dollars rises in the London market and falls in the New York market. Ultimately, it will equal in both the markets and arbitrage comes to an end. If the exchange rate between the dollar and the pound rises to $ 2.60 = £ 1 through time, the dollar is said to depreciate with respect to the pound, because now more dollars are needed to buy one pound. When the rate of exchange between the dollar and the pound falls to $ 2.40 = £ 1, the value of the dollar is said to appreciate because now less dollars are required to purchase one pound. If the value of the first currency depreciates that of the other appreciates, and vice versa. Thus a depreciation of the dollar against the pound is the same thing as the appreciation of the pound against the dollar, and vice versa.

14.2 Components of Exchange Rate

Forward and Spot Exchange Rates

There is much empirical work on forward foreign exchange rates as predictors of future spot exchange rates. There is also a growing literature on whether forward rates contain variation in premiums. There is a general concensus that forward rates have little if any power to forecast changes in spot rates. There is less consensus on the existence of time varying premiums in forward rates. Frankel (1982) and Domowitz and Hakkio (1983) fail to identify such premiums, while Hsieh (1982). Hansen and Hodrick (1983), Hodrick and Srivastava (1984), and Korajczyk (1983) find evidence consistent with time varying premiums.

This unit tests a model for joint measurement of variation in the premium and expected future spot rate components of forward rates. Conditional on the hypothesis that the forward market is efficient or rational, we find reliable evidence that both components of forward rates vary through time. More startling are the conclusions that (a) most of the variation in forward rates is variation in premiums, and (b) the premium and expected future spot rate components of forward rates are negatively correlated.

Theoretical framework

The forward exchange rate f_t observed at time t for an exchange at $t+1$ is the market determined certainty equivalent of the future spot exchange rate s_{t+1}. One way to split this certainty equivalent into an expected future spot rate and a premium is

$$F_t = E(S_{t+1}) + P_t \qquad \qquad \dots (1)$$

where $F_t = ln\ f_t$, $S_{t+1} = ln\ s_{t+1}$, and the expected future spot rate, $E(S_{t+1})$, is the rational or efficient forecast, conditional on all information available at t. Logs are used (a) to make the analysis independent of whether exchange rates are expressed as units of currency i per unit of currency j or units of j per unit of i, and (b) because some models for the premium can be stated in logs.

Eq. (1) is no more than a particular definition of the premium component of the forward rate. To give the equation economic content, a model that describes the determination of P, is required. Examples of such models are discussed later. For the statistical analysis of the premium and expected future spot rate components of the forward rate, however, it suffices that the forward rate is the market determined certainty equivalent of the future spot rate.

Statistics

From (1) the difference between the forward rate and the current spot rate is

$$F_t - S_t = P_1 + E(S_{t+1} - S_t). \qquad \qquad \dots (2)$$

Consider the regressions of $F_t - S_{t+1}$ and $S_{t+1} - S_t$ (both observed at $t+1$) on $F_t - S_t$ (observed at t),

$$F_t - S_{t+1} = \alpha_1 + \beta_1(F_t - S_t) + \varepsilon_{1,t+1}, \qquad \qquad \dots (3)$$

$$S_{t+1} - S_t = \alpha_2 + \beta_2(F_t - S_t) + \varepsilon_{2,t+1}. \qquad \qquad \dots (4)$$

Estimates of (4) tell us whether the current forward-spot differential, $F_t - S_t$ has power to predict the future change in the spot rate, $S_{t+1} - S_t$ Evidence that β_2 is reliably non-zero means that the forward rate observed at t has information about the spot rate to be observed at $t+1$. Likewise, since $F_t - S_{t+1}$ is the premium P_t plus $E(S_{t+1}) - S_{t+1}$, the random error of the rational forecast $E(S_{t+1})$, evidence that β_1 in (3) is reliably non-zero means that the premium component of $F_t - S_t$ has variation that shows up reliably in $F_t - S_{t+1}$.

With the assumption that the expected future spot rate in the forward rate is efficient or rational, the regression coefficients in (3) and (4) are

$$\beta_1 = \frac{cov(F_t - S_{t+1}, F_t - S_t)}{\sigma^2(F_t - S_t)},$$

$$= \frac{\sigma^2(P_t) + cov(P_t, E(S_{t+1} - S_t))}{\sigma^2(P_t) + \sigma^2(E(S_{t+1} - S_t)) + 2\,cov(P_t, E(S_{t+1} - S_t))}, \qquad \dots (5)$$

$$\beta_2 = \frac{cov(S_{t+1} - S_t, F_t - S_t)}{\sigma^2(F_t - S_t)}$$

$$= \frac{\sigma^2(E(S_{t+1} - S_t)) + cov(P_t, E(S_{t+1} - S_t))}{\sigma^2(P_t) + \sigma^2(E(S_{t+1} - S_t)) + 2\,cov(P_t, E(S_{t+1} - S_t))} \qquad \dots (6)$$

In the special case where P_t and $E(S_{t+1} - S_t)$ are uncorrected, the regression coefficients β_1 and β_2 split the variance of $F_t - S_t$ into two parts : the proportion due to the variance of the premium and the proportion due to the variance of the expected change in the spot rate. When the two components of $F_t - S_t$ are correlated, the contribution of covariation between P_t and $E(S_{t+1} - S_t)$ to $\sigma^2(F_t - S_t)$ is divided equally between β_1, and β_2. The regression coefficients still include the proportions of $\sigma^2(F_t - S_t)$ due to $\sigma^2(P_t)$ and $\sigma^2(E(S_{t+1} - S_t))$, but the simple interpretation of β_1 and β_2 obtained when P_t and $E(S_{t+1} - S_t)$ are uncorrected is lost. The troublesome $\text{cov}(P_t, E(S_{t+1} - S_t))$ in (5) and (6) is a central issue in the empirical tests.

Since $F_t - S_{t+1}$ and $S_{t+1} - S_t$, sum to $F_t - S_t$, the sum of the intercepts in (3) and (4) must be zero, the sum of the slopes must be 1.0, and the disturbances, period-by-period, must sum to 0.0. In other words, regressions (3) and (4) contain identical information about the variation of the P_t and $E(S_{t+1} - S_t)$ components of $F_t - S_t$, and in principle there is no need to show both. I contend, however, that joint analysis of the regressions is what makes clear the information that either contains. Thus, regression (4) of the change in the spot rate, $S_{t+1} - S_t$, on the forward rate minus the current spot rate, $F_t - S_t$, is common in the literature. It is also widely recognized that deviations of β_2 in (4) from 1.0 can somehow be due to a time varying premium in the forward rate. To my knowledge, however, the explicit interpretation of the regression coefficients provided by (5) and (6) is not well known. In particular, it is not widely recognized that, given an efficient or rational exchange market, the deviation of β_2 from 1.0 is a direct measure of the variation of the premium in the forward rate. The complementarity of the regression coefficients in (3) and (4) which is described in (5) and (6) helps us to interpret some of the anomalous results observed for estimates of (4).

Economics

Since a major conclusion of the empirical work is that variation in forward rates is mostly variation in premiums, some discussion of the economics of premiums is warranted. Using more precise notation, let f_t^{ij} and s_t^{ij} be the forward and spot exchange rates (units of currency *i* per unit of currency *j*) observed at *t*, and let R_{it} and R_{jt} be the nominal interest rates observed at *t* on discount bonds denominated in currencies *i* and *j*. The bonds have either zero or identical default risks, and they have the same maturity as f_t^{ij}.

With open international bond markets, the no arbitrage condition of interest rate parity (IRP) implies

$$f_t^{ij} / s_t^{ij} = (1 + R_{it}) / (1 + R_{jt}). \qquad \text{... (7)}$$

Thus, the difference between the forward and spot exchange rates observed at *t* is directly related to the difference between the interest rates on nominal bonds denominated in the two currencies. Any premium in the forward rate must be explainable in terms of the interest rate differential.

For example (and keep in mind that it is just an example), suppose (a) that exchanges rates are characterized by complete purchasing power parity (PPP), and (b) that the Fisher equation holds for nominal interest rates. Let V_{it} and V_{jt} be the price levels in the two countries, let $\Delta_{i,t+1} = ln(V_{i,t+1} / V_{it})$ and $\Delta_{j,t+1} = ln(V_{j,t+1} / V_{jt})$ be their inflation rates, and let $r_{i,t+1}$ and $r_{j,t+1}$ be the *ex post* continuously compounded real returns on their nominal bonds. Taking logs in (7) and applying the Fisher equation to the resulting continuously compounded nominal interest rates, we have

$$F_t^{ij} - S_t^{ij} = \left[E\left(r_{i,t+1}\right) + E\left(\Delta_{i,t+1}\right)\right] - \left[E\left(r_{j,t+1}\right) + E\left(\Delta_{i,t+1}\right)\right]$$

$$= \left[E\left(r_{i,t+1}\right) - E\left(r_{j,t+1}\right)\right] + \left[E\left(ln V_{i,t+1}\right) - E\left(ln V_{j,t+1}\right)\right]$$

$$- \left[ln V_{it} - ln V_{jt}\right] \qquad\qquad \text{... (8)}$$

With complete PPP, $s_t^{ij} = V_{it} / V_{jt}$, that is, the spot exchange rate is the ratio of the price levels in the two countries, and (11) reduces to

$$F_t^{ij} = \left[E\left(r_{i,t+1}\right) - E\left(r_{j,t+1}\right)\right] + E\left(S_{t+1}^{ij}\right) \qquad\qquad \text{... (9)}$$

In words, with the Fisher equation, interest rate parity and purchasing power parity, the premium P_t in the forward rate expression (1) is just the difference between the expected real returns on the nominal bonds of the two countries. Thus, the variables that determine the difference between the expected real returns on the nominal bonds (for example, differential purchasing power risks of their nominal payoffs) also explain the premium in the forward rate. This interpretation applies to any model of international capital market equilibrium characterized by IRP, PPP, and the Fisher equation for nominal interest rates. Examples are the international version of the Sharpe (1964) and Lintner (1965) model discussed by Fama and Farber (1979) or the version of the Lucas (1978) model discussed by Hodrick and Srivastava (1984).

The lock between the premium in the forward exchange rate and the interest rates on the nominal bonds of two countries is the direct consequence of the interest rate parity condition (7) of an open international bond market. For example, using IRP and an international version of the Breeden (1979) model, Stulz (1981) derives an expression for the forward rate similar to (1) or (9), but for a world in which (a) complete PPP does not hold, and (b) differential tastes for consumption goods combine with uncertainty about relative prices to strip the Fisher equation of its meaning.

Data and summary statistics

Spot exchange rates and thirty-day forward rates for nine major currencies are taken from the Harris Bank Data Base supported by the Center for Studies in International Finance of the University of Chicago. The rates are Friday closes sampled at four-week intervals. There are 122 observations covering the period August 31, 1973, to December 10,1982. All rates are U.S. dollars per unit of foreign currency.

Table 1 shows means, standard deviations, and autocorrelations of $S_{t+1} - S_t$ (the four-week change in the spot rate), $F_t - S_{t+1}$ (the thirty-day forward rate minus the spot rate observed four weeks later), and $F_t - S_t$ (the forward rate minus the current spot rate). Since the forward and spot rates are in logs and the differences are multiplied by 100, the three variables are on a percent per month basis.

The standard deviations of $F_t - S_{t+1}$ in table 1 are larger than the standard deviations of $S_{t+1} - S_t$. Thus, in terms of standard deviation of forecast errors, the current spot rate is a better predictor of the future spot rate than the current forward rate. However, variation in the premium component of the forward rate can obscure the power of the prediction of the future spot rate in the forward rate. This is the problem that the complementary regressions (3) and (4) are meant to alleviate.

Consistent with the previous literature, the autocorrelations of changes in spot rates, $S_{t+1} - S_t$, are close to zero. Thus, if the expected component of the changes, $E(S_{t+1} - S_t)$, varies in an autocorrelated way, this is not evident in the behavior of the observed changes. The $F_t - S_{t+1}$ for different countries also show little autocorrelation. $F_t - S_{t+1}$ is the premium, P_t, plus the forecast error, $E(S_{t+1}) - S_{t+1}$,

Table 1: Autocorrelations, means, and standard deviations : 8/31/73-12/10/82, N = 122.

Country	ρ	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_0	ρ_1	ρ_2	Mean	Std. dev.
							$S_{t+1} - S_t$							
Belgium	0.05	0.08	0.07	− 0.03	0.02	− 0.03	− 0.01	0.06	0.14	− 0.07	0.20	0.01	− 0.25	3.11
Canada	0.13	− 0.24	0.08	0.05	0.03	0.00	− 0.16	0.06	0.07	− 0.15	0.16	0.03	− 0.17	1.12
France	− 0.04	0.06	0.14	− 0.02	0.14	0.04	0.01	− 0.02	0.10	− 0.12	0.11	− 0.01	− 0.43	3.01
Italy	0.01	0.15	− 0.03	− 0.11	0.09	− 0.01	0.09	0.04	0.14	− 0.17	0.12	− 0.01	− 0.73	2.80
Japan	016	− 0.11	0.03	0.13	0.15	− 0.09	− 0.04	0.07	0.05	− 0.09	− 0.10	0.11	0.07	3.05
Netherlands	0.02	0.05	0.05	− 0.14	− 0.01	− 0.01	− 0.01	0.04	0.04	− 0.06	0.17	− 0.02	− 0.04	3.01
Switzerland	0.01	0.08	0.03	− 0.11	0.09	0.01	− 0.05	− 0.08	0.01	− 0.04	0.07	− 0.01	0.26	3.76
United Kingdom	0.15	0.04	0.10	− 0.07	0.09	0.04	0.16	0.01	0.02	0.07	0.05	0.14	− 0.36	2.58
West Germany	0.01	0.08	0.01	− 0.13	0.00	− 0.04	0.02	0.06	0.07	− 0.05	0.17	− 0.00	− 0.03	3.08
							$F_t - S_{t+1}$							
Belgium	0.11	0.10	0.08	− 0.02	0.02	− 0.05	− 0.03	0.04	0.11	− 0.08	0.19	0.00	0.09	3.22
Canada	0.17	− 0.21	0.07	0.04	0.01	− 0.01	− 0.16	0.04	0.06	− 0.14	0.15	0.02	0.08	1.16
France	0.01	0.09	0.16	− 0.01	0.14	0.04	0.00	− 0.02	0.06	− 0.10	0.12	− 0.01	0.17	3.10
Italy	0.08	0.17	− 0.02	− 0.11	0.07	− 0.03	0.08	0.03	0.11	− 0.17	0.07	− 0.03	− 0.07	2.95
Japan	0.21	− 0.05	0.07	0.13	0.14	− 0.09	− 0.05	0.05	0.05	− 0.07	− 0.08	0.12	0.10	3.15
Netherlands	0.07	0.08	0.07	− 0.11	0.01	− 0.00	− 0.00	0.04	0.04	− 0.05	0.19	0.01	0.21	3.08
Switzerland	0.05	0.10	0.06	− 0.09	0.09	0.01	− 0.04	− 0.08	0.01	− 0.02	0.08	0.01	0.23	3.82
United Kingdom	0.19	0.08	0.12	− 0.06	0.09	0.03	0.15	0.01	0.01	0.05	0.07	0.13	0.13	2.65
West Germany	0.03	0.09	0.03	− 0.12	0.00	− 0.04	0.02	0.05	0.07	− 0.04	0.18	0.02	0.33	3.12

Autocorrelations

Notes

						$F_t - S_t$							
Belgium	0.67	0.44	0.33	0.33	0.18	0.04	– 0.01	0.01	0.00	– 0.01	0.00	– 0.01	– 0.16	0.41
Canada	0.82	0.63	0.49	0.38	0.36	0.32	0.33	0.36	036	0.35	0.26	0.23	– 0.09	0.17
France	0.65	0.45	0.36	0.28	0.24	0.18	0.17	0.18	0.23	0.28	0.26	0.23	– 0.25	0.44
Italy	0.68	0.47	0.37	0.28	0.23	0.19	0.18	0.19	0.19	0.14	0.03	0.04	– 0.80	0.66
Japan	0.85	0.69	0.61	0.47	0.34	0.30	0.22	0.16	0.21	0.24	0.22	0.20	0.17	0.64
Netherlands	0.72	0.55	0.40	0.28	0.24	0.20	0.20	0.23	0.22	0.21	0.15	0.09	0.17	0.32
Switzerland	0.86	0.73	0.61	0.52	0.47	0.46	0.48	0.49	0.50	0.49	0.44	0.40	0.48	0.37
United Kingdom	0.87	0.75	0.64	0.51	0.43	0.36	0.31	0.28	0.25	0.20	0.16	0.15	– 0.23	0.35
West Germany	0.78	0.56	0.39	0.26	0.20	0.20	0.26	0.34	0.42	0.46	0.40	0.33	0.30	0.24

*All exchange rates are U.S. dollars per unit of foreign currency. $S_{t+1} - S_t$, is the four-week change in the spot exchange rate; $F_t - S_{t+1}$ is the thirty-day forward rate minus the spot rate observed four weeks later, $F_t - S_t$, is the forward rate minus the current spot rate. The means and standard deviations of the variables are on a percent per month basis. Under the hypothesis that the true autocorrelations are 0.0, the standard error of the sample autocorrelations is about 0.09.

which should be white noise. Thus, any autocorrelation of the premium is not evident in the time series behavior of $F_t - S_{t+1}$.

The autocorrelations of $F_t - S_t$ tell a different story. The first-order autocorrelations are 0.6S or greater, and the decay of the autocorrelations at successive lags suggests a first-order autoregressive process. This is confirmed by the partial autocorrelations (not shown) which are large at lag 1 but close to zero at higher-order lags. Since $F_t - S_t$ is the premium, P_t, plus the expected change in the spot rate, $E(S_{t+1} - S_t)$, the autocorrelations of $F_t - S_t$ indicate that P_t and/or $E(S_{t+1} - S_t)$ vary in an autocorrelated way.

The difference between the behavior of the autocorrelations of $F_t - S_t$ and those of $E(S_{t+1} - S_t)$ and $F_t - S_{t+1}$ is easily explained. The standard deviations of $F_t - S_t$ are between 0.17 and 0.66 percent per month, whereas those of either $S_{t+1} - S_t$ or $F_t - S_{t+1}$ are typically greater than 3.0 percent per month. Thus, the autocorrelation of P_t and/or $E(S_{t+1} - S_t)$, which shows up in the time series behavior of $F_t - S_t$, is buried in the high variability of the unexpected components of $F_t - S_{t+1}$ and $S_{t+1} - S_t$.

Regression tests

OLS estimates

Table 2 shows the estimated regressions of $F_t - S_{t+1}$ and $S_{t+1} - S_t$ on $F_t - S_t$. Only one set of coefficient standard errors, residual standard errors and residual autocorrelations is shown for each country. This reflects the complementarity of the $F_t - S_{t+1}$ and $S_{t+1} - S_t$ regressions for each country. The intercept estimates in the two regressions sum to zero, the slope coefficients sum to one, and the sum of the two residuals is zero on a period-by-period basis.

Since the regressor $F_t - S_t$ has low variation relative to $F_t - S_{t+1}$ and $S_{t+1} - S_t$, the coefficients of determination (R_1^2 and R_2^2) for the regressions are small, and they are smaller for the $S_{t+1} - S_t$ regressions than for the $F_t - S_{t+1}$ regressions. The regression residuals, like the dependent variables, show little autocorrelation.

The anomalous numbers in table 2 are the estimates of the regression slope coefficients, β_1 and β_2 According to (5) and (6), the slope coefficient in the regression of $F_t - S_{t+1}$ on $F_t - S_t$ contains the proportion of the variance of $F_t - S_t$ due to variation in its premium component, P_t, while the slope coefficient in the regression of $S_{t+1} - S_t$ on $F_t - S_t$ contains the proportion of the variance of $F_t - S_t$ due to variation in the expected change in the spot rate, $E(S_{t+1} - S_t)$. The coefficients clearly cannot be interpreted in terms of these

proportions alone, since the coefficients in the $S_{t+1} - S_t$ regressions are always negative so that those in the $F_t - S_{t+1}$, regressions are greater than 1.0.

Inspection of (5) and (6) indicates an explanation for the strange estimates of β_1 and β_2. Since $\sigma^2(E(S_{t+1} - S_t))$ in (6) must be non-negative, a negative estimate of β_2 implies that

Notes

Table 2: OLS regressions : 8/31/73-12/10/82. N = 122.*

$$F_t - S_{t+1} = \hat{\alpha}_1 + \hat{\beta}_1(F_t - S_t) + \hat{\varepsilon}_{1,t+1}. \qquad S_{t+1} - S_t = \hat{\alpha}_2 + \hat{\beta}_2(F_t - S_t) + \hat{\varepsilon}_{2,t+1}.$$

| Country | $\hat{\alpha}_1$ | $\hat{\beta}_1$ | $\hat{\alpha}_2$ | $\hat{\beta}_2$ | $s(\hat{\alpha})$ | $s(\hat{\beta})$ | R_1^2 | R_2^2 | $s(\hat{\varepsilon})$ | Residual autocorrelations | | | | | |
										ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
Belgium	0.50	2.58	– 0.50	– 1.58	0.30	0.68	0.11	0.04	3.05	0.01	0.06	0.06	– 0.03	0.02	0.02
Canada	0.25	1.87	– 0.25	– 0.87	0.11	0.61	0.07	0.01	1.12	0.12	– 0.23	0.10	0.07	0.06	0.03
France	0.64	1.87	– 0.64	– 0.87	0.31	0.63	0.07	0.01	3.00	– 0.07	0.04	0.13	– 0.03	0.15	0.04
Italy	1.14	1.51	– 1.14	– 0.51	0.40	0.38	0.11	0.01	2.79	– 0.00	0.16	– 0.01	– 0.09	0.10	0.01
Japan	– 0.12	1.29	0.12	– 0.29	0.29	0.43	0.07	0.00	3.06	0.15	– 0.12	0.03	0.13	0.16	– 0.08
Netherlands	– 0.21	2.43	0.21	– 1.43	0.31	0.86	0.06	0.01	2.99	– 0.03	0.03	0.02	– 0.17	– 0.01	– 0.02
Switzerland	– 0.81	2.14	0.81	– 1.14	0.56	0.92	0.04	0.00	3.75	– 0.02	0.06	0.01	– 0.12	0.10	0.02
United Kingdom	0.57	1.90	– 0.57	– 0.90	0.28	0.66	0.06	0.01	2.57	0.13	0.03	0.11	– 0.06	0.10	0.05
West Germany	– 0.36	2.32	0.36	– 1.32	0.44	1.15	0.03	0.00	3.08	– 0.01	0.07	0.00	– 0.13	0.01	– 0.03

*R_1^2 and R_2^2 are the coefficients of determination (regression R^2) for the $F_t - S_{t+1}$ and $S_{t+1} - S_t$, regressions. The complete complementarity of the $F_t - S_{t+1}$ and $S_{t+1} - S_t$ regressions for each country means that the standard errors $s(\hat{\alpha})$ and $s(\hat{\beta})$ of the estimated regression coefficients, the residual standard error $s(\hat{\varepsilon})$, and the residual autocorrelations, ρ_t, are the same for the two regressions. Under the hypothesis that the true autocorrelations are zero, the standard error of the estimated residual autocorrelations is about 0.09.

$\mathrm{cov}\left(P_t, E(S_{t+1} - S_t)\right)$ is negative and larger in magnitude than $\sigma^2\left(E(S_{t+1} - S_t)\right)$. The complementary estimate of $\beta_1 > 1$ then implies that $\mathrm{cov}\left(P_t, E(S_{t+1} - S_t)\right)$ is smaller in absolute magnitude than $\sigma^2(P_t)$, and thus that $\sigma^2(P_t)$ is larger than $\sigma^2\left(E(S_{t+1} - S_t)\right)$.

The non-zero covariance between P_t and $E(S_{t+1} - S_t)$ prevents us from using the regression coefficients to estimate the levels of $\sigma^2(P_t)$ and $\sigma^2\left(E(S_{t+1} - S_t)\right)$. With (5) and (6), however, we can estimate the difference between the two variances as a proportion of $\sigma^2(F_t - S_t)$.

$$\beta_1 - \beta_2 \;=\; \frac{\sigma^2(P_t) - \sigma^2\left(E(S_{t+1} - S_t)\right)}{\sigma^2(F_t - S_t)}. \qquad \ldots (10)$$

The differences between the estimates of β_1 and β_2 in table 2 range from 1.58 (Japan) to 4.16 (Belgium). Except for Japan, all the differences between the estimated coefficients are greater than 2.0. Thus, the point estimates are that the difference between the variance of the premium, P_t, and the variance of the expected change in the spot rate, $E(S_{t+1} - S_t)$, in $F_t - S_t$ is typically more than twice the variance of $F_t - S_t$. Moreover, since β_1 and β_2 sum to 1.0, the estimates of the regression coefficients are perfectly negatively correlated, and the standard error of their difference is twice their common standard error. Only the estimates of $\beta_1 - \beta_2$ for Japan, Switzerland, and West Germany are less than two standard errors from zero, and all are more than 1.5 standard errors from zero. Thus, we can conclude that $\sigma^2(P_t)$ is reliably greater than $\sigma^2\left(E(S_{t+1} - S_t)\right)$.

In short, negative covariation between P_t and $E(S_{t+1} - S_t)$ attenuates the variability of $F_t - S_t$, and obscures the interpretation of the regression slope coefficients in (3) and (4). Nevertheless the regression slope coefficients provide the interesting information that both the premium, P_t, and the expected change in the spot rate, $E(S_{t+1} - S_t)$, in $F_t - S_t$, vary through time, and $\sigma^2(P_t)$ is large relative to $\sigma^2\left(E(S_{t+1} - S_t)\right)$.

A good story for negative covariation between P_t and $E(S_{t+1} - S_t)$ is difficult to tell. For example, in the PPP model for the exchange rate underlying (9), the dollar is expected to appreciate relative to a foreign currency, that is, $E(S_{t+1} - S_t)$ is negative, when the expected inflation rate in the U.S. is lower than in the foreign country. (Remember that the exchange rates are all expressed as dollars per unit of foreign currency.) A negative $\mathrm{cov}\left(P_t, E(S_{t+1} - S_t)\right)$ then implies a higher purchasing power risk premium in the expected real returns on dollar denominated bonds relative to foreign currency bonds when the anticipated U.S. inflation rate is low relative to the anticipated foreign inflation rate. We return to economic interpretations of the negative covariance between the P_t and E $(S_{t+1} - S_t)$ components of $F_t - S_t$ after exploring some purely statistical possibilities.

SUR estimates

The apparent negative covariation between P_t and E $(S_{t+1} - S_t)$ may be sampling error. All the slope

coefficients in the $F_t - S_{t+1}$ regressions are more than two standard errors above 0.0, but only one (Belgium) is more than two standard errors above 1.0. Equivalently, only one of the negative slope coefficients in the $S_{t+1} - S_t$ regressions (Belgium) is more than two standard errors below zero. Perhaps the appropriate conclusion is that all variation through time in $F_t - S_t$ is variation in premiums, and there is no variation in expected changes in spot rates.

Individually testing the β_1 coefficients in table 2 against 1.0 (or the β_2 coefficients against 0.0) does not provide the appropriate joint test that all $\beta_1 = 1.0$ (or all $\beta_2 = 0$). An appropriate joint test takes into account the high correlation of $F_t - S_{t+1}$ (or $S_{t+1} - S_t$) across currencies, documented in table 3. Such cross-correlation is to be expected given that (a) all exchange rates are measured relative to the U.S. dollar, and (b) most of the European countries are involved in attempts to control the movements of their exchange rates relative to one another during the sample period. Table 3 also indicates that, with the possible exception of Canada, the correlations of the regressor variable $F_t - S_t$ across countries are generally lower than the correlations of $S_{t+1} - S_t$ or $F_t - S_{t+1}$ across the countries. Thus, there is reason to suspect that joint estimation of the $F_t - S_{t+1}$ (or the $S_{t+1} - S_t$) regressions for different countries will improve the precision of the coefficient estimates.

The coefficient estimates obtained when Zellner's (1962) 'seemingly unrelated regression' (SUR) technique is used to estimate either the $F_t - S_{t+1}$ regressions for different countries or the $S_{t+1} - S_t$ regressions are summarized in part A of table 4. As anticipated, joint estimation substantially improves the precision of the estimated slope coefficients. The $s(\beta)$ in table 4 are often less than half those for the OLS estimates in table 2. Moreover, the slope coefficients in the SUR versions of the $S_{t+1} - S_t$ regressions are generally closer to zero than in the OLS regressions which means that the coefficients in the complementary $F_t - S_{t+1}$ regressions are generally closer to 1.0. (Canada and Switzerland are exceptions.)

Table 4 also shows F tests on various joint hypotheses on the coefficients. The hypothesis that all the slope coefficients β_2 in the $S_{t+1} - S_t$ regressions (or all the slope coefficients β_1 in the $F_t - S_{t+1}$ regressions) are equal is consistent with the data. However, the hypothesis that all $\beta_2 = 0.0$ (or all $\beta_1 = 1.0$) yields

Table 3: Correlations of $S_{t+1} - S_t$, $F - S_{t+1}$ and $F_t - S_t$ across countries.

	Belgium	Canada	France	Italy	Japan	Netherlands	Switzer- land	United Kingdom	West Germany
				$S_{t+1} - S_t$					
Belgium	1.00								
Canada	0.19	1.00							
France	0.84	0.18	1.00						
Italy	0.69	0.10	0.77	1.00					
Japan	0.52	0.04	0.56	0.48	1.00				
Netherlands	0.94	0.19	0.85	0.72	0.50	1.00			
Switzerland	0.81	0.13	0.76	0.64	0.53	0.81	1.00		
United Kingdom	0.57	0.18	0.54	0.54	0.46	0.56	0.51	1.00	
West Germany	0.94	0.17	0.84	0.71	0.54	0.96	0.85	0.53	1.00
				$F_t - S_{t+1}$					
Belgium	1.00								
Canada	0.20	1.00							
France	0.85	0.16	1.00						

Italy	0.69	0.07	0.74	1.00					
Japan	0.51	0.07	0.55	0.46	1.00				
Netherlands	0.94	0.20	0.85	0.71	0.51	1.00			
Switzerland	0.80	0.14	0.76	0.61	0.55	0.80	1.00		
United Kingdom	0.55	0.16	0.51	0.51	0.47	0.56	0.50	1.00	
West Germany	0.93	0.18	0.84	0.69	0.56	0.96	0.85	0.53	1.00

$$F_t - S_t$$

Belgium	1.00								
Canada	0.55	1.00							
France	0.57	0.43	1.00						
Italy	0.59	0.43	0.51	1.00					
Japan	0.31	0.12	0.36	0.32	1.00				
Netherlands	0.54	0.40	0.31	0.49	0.49	1.00			
Switzerland	0.38	0.25	0.51	0.43	0.77	0.59	1.00		
United Kingdom	0.45	0.36	0.31	0.47	0.65	0.74	0.63	1.00	
West Germany	0.49	0.50	0.52	0.49	0.73	0.69	0.88	0.72	1.00

[a]$S_{t+1} - S_t$ is the four-week change in the spot exchange rate : $F_t - S_{t+1}$ is the thirty-day forward rate minus the spot rate observed four weeks later; $F_t - S_t$ is the forward rate minus the current spot rate.

Table 4: SUR regressions : 8/31/73–12/10/82, N = 122.[a]

$$F_t - S_{t+1} = \hat{\alpha}_1 + \hat{\beta}_1(F_t - S_t) + \hat{\varepsilon}_{1,t+1}, \quad S_{t+1} - S_t = \hat{\alpha}_2 + \hat{\beta}_2(F_t - S_t) + \hat{\varepsilon}_{2,t+1}.$$

Part A : Unconstrained

Country	$\hat{\alpha}_2(=-\hat{\alpha}_1)$	$\hat{\beta}_2(=1-\hat{\beta}_1)$	$s(\hat{\alpha})$	$s(\hat{\beta})$
Belgium	– 0.36	– 0.72	0.28	0.24
Canada	– 0.26	– 1.04	0.11	0.59
France	– 0.48	– 0.21	0.28	0.30
Italy	– 1.08	– 0.44	0.32	0.24
Japan	0.12	– 0.28	0.28	0.35
Netherlands	0.10	– 0.78	0.27	0.25
Switzerland	0.81	– 1.15	0.42	0.50
United Kingdom	– 0.52	– 0.69	0.26	0.51
West Germany	0.23	– 0.89	0.29	0.32
F tests	1. All β_2 (or β_1) equal	F = 0.73	P level = 0.66	
	2. All α_2 (or α_1) equal	F = 5.14	P level = 0.0001	
	3. All $\beta_2 = 0.0$ (or $\beta_1 = 1.0$)	F = 2.81	P level = 0.003	

Notes

	Part B : Constrained									

$$S_{t+1} - S_t = \hat{\alpha}_B + \hat{\alpha}_C + \hat{\alpha}_F + \hat{\alpha}_I + \hat{\alpha}_J + \hat{\alpha}_N + \hat{\alpha}_S + \hat{\alpha}_{UK} + \hat{\alpha}_{WG} + \beta_2 \ (F_t - S_t)$$

	$\hat{\alpha}_B$	$\hat{\alpha}_C$	$\hat{\alpha}_F$	$\hat{\alpha}_I$	$\hat{\alpha}_J$	$\hat{\alpha}_N$	$\hat{\alpha}_S$	$\hat{\alpha}_{UK}$	$\hat{\alpha}_{WG}$	β_2
	– 0.34	– 0.22	– 0.57	– 1.20	0.17	0.07	0.54	– 0.49	0.14	– 0.58
	(0.28)	(0.10)	(0.27)	(0.27)	(0.28)	(0.27)	(0.35)	(0.23)	(0.28)	(0.13)

F test All α equal F = 5.68 P level = 0.0001

[a]Like the OLS regressions in table 2, the SUR regressions are completely complementary; that is, the intercepts in the $F_t - S_{t+1}$ and $S_{t+1} - S_t$ regressions sum to 0.0, the slopes sum to 1.0, and the residuals sum to 0.0 period-by-period. The subscripts on the $\hat{\alpha}$ in the constrained $S_{t+1} - S_t$ regressions indicate countries.

a test statistic far out in the tail of the F distribution (beyond the 0.997 fractile) which suggests rejection of the hypothesis. Thus, we are left with the uncomfortable conclusion that the negative estimates of β_2 in the regressions of $S_{t+1} - S_t$ on $F_t - S_t$ are the result of negative covariation between the P_t and $E(S_{t+1} - S_t)$ components of $F_t - S_t$.

Finally, since the hypothesis that the slope coefficients in the $S_{t+1} - S_t$ (or $F_t - S_{t+1}$) regressions are equal across countries is consistent with the data, we can use the SUR technique to estimate the regressions subject to the equality constraint. The results for the $S_{t+1} - S_t$ regressions are shown in part B of table 4. For all but three countries (France, Italy and Japan) the constrained estimate of β_2, – 0.58, is closer to 0.0 than the unconstrained estimate in part A of the table. However, constraining the estimate of β_2 to be equal across countries so lowers the standard error of the estimate that β_2 is now more than four standard errors from 0.0.

Subperiod results

Some argue that the nature of the flexible exchange rate system during our sample period is not well understood by market participants until the late 1970's. [See, for example, Hansen and Hodrick (1983).] Thus, the properties of forward exchange rates as predictors of future spot rates may be different during later years. To check on this possibility, the tests in tables 1 to 4 are replicated for the two 61-month subperiods covered by the data. The results are summarized in tables 5 to 7. The subperiod results also help to alleviate any statistical problems caused by changes in variance during the sample period.

There are some differences between the two subperiods. For example, the summary statistics of table 5 document an increase in the variability of $S_{t+1} - S_t$ and $F_t - S_{t+1}$ for the later period. There is no corresponding increase in the variability of $F_t - S_t$. The implied conclusion is that the higher variability of $S_{t+1} - S_t$ and $F_t - S_{t+1}$ in later years reflects increased uncertainty about the *ex post* change in the spot rate with no corresponding increase in the variability of the *ex ante* E $(S_{t+1} - S_t)$ and P_t components of $F_t - S_t$.

The mean values of the variables do not suggest improved market forecasts of future spot rates during the later subperiod. The mean of $F_t - S_t$ more often has the same sign as the mean of $S_{t+1} - S_t$ during the earlier subperiod (seven of nine versus five of nine for the later period). Moreover, although the dollar appreciates relative to all nine currencies during the later period (the means of $S_{t+1} - S_t$ are all negative), all the means of $F_t - S_t$ move upward. Thus, either the forward rate on average becomes a less rational predictor of the future spot rate during the later period, or, as suggested by the regression results, there is opposite movement in the premium component of $F_t - S_t$ which more than offsets movement in the expected change in the spot rate.

Table 5: OLS regressions for 61-month subperiods.[a]

$$F_t - S_{t+1} = \hat{\alpha}_1 + \hat{\beta}_1(F_t - S_t) + \hat{\varepsilon}_{1,t+1}, \; S_{t+1} - S_t = \hat{\alpha}_2 + \hat{\beta}_2(F_t - S_t) + \hat{\varepsilon}_{2,t+1}$$

Country	$\hat{\alpha}_2(=-\hat{\alpha}_1)$	$\hat{\beta}_2(=1-\hat{\beta}_2)$	$s(\hat{\alpha})$	$s(\hat{\beta})$	R_1^2	R_2^2	$s(\hat{\varepsilon})$	ρ_1
First subperiod : 8/31/ 73-4/ 7/ 78								
Belgium	− 0.20	− 1.42	0.40	0.83	0.13	0.05	2.55	0.05
Canada	− 0.25	− 0.32	0.16	0.77	0.05	0.00	1.01	0.19
France	− 0.79	− 1.38	0.51	0.87	0.11	0.04	2.48	0.06
Italy	− 1.17	− 0.51	0.60	0.47	0.15	0.02	2.58	0.17
Japan	0.37	0.31	0.29	0.42	0.04	0.01	2.18	0.20
Netherlands	0.31	− 1.22	0.34	1.14	0.06	0.02	2.68	0.04
Switzerland	0.52	0.81	0.47	1.40	0.00	0.01	2.88	0.10
United Kingdom	− 0.47	0.02	0.52	1.04	0.02	0.00	2.35	0.14
West Germany	0.62	− 2.60	0.45	2.12	0.05	0.03	2.65	0.14
Second subperiod : 5/5/78-12/10/82								
Belgium	− 0.74	− 1.32	0.45	1.18	0.06	0.02	3.50	− 0.02
Canada	− 0.23	− 1.64	0.17	0.98	0.11	0.05	1.22	0.06
France	− 0.70	− 0.22	0.45	1.11	0.02	0.00	3.47	− 0.12
Italy	− 1.15	− 0.60	0.58	0.80	0.06	0.01	3.04	− 0.12
Japan	0.82	− 1.84	0.92	1.46	0.06	0.03	3.72	0.11
Netherlands	0.02	− 1.18	0.73	1.77	0.03	0.01	3.32	− 0.06
Switzerland	1.66	− 2.44	1.98	2.50	0.03	0.02	4.47	− 0.06
United Kingdom	− 0.36	− 2.83	0.35	1.12	0.17	0.10	2.71	− 0.03
West Germany	− 0.30	− 0.04	1.05	2.10	0.00	0.00	3.48	− 0.09

[a] R_1^2 and R_2^2 are the coefficients of determination (regression R^2) for the $F_t - S_{t-1}$ and $S_{t+1} - S_t$ regressions. The complete complementarity of the $F_t - S_{t+1}$ and $S_{t+1} - S_t$ regressions for each country means that the standard errors $s(\hat{\alpha})$ and $s(\hat{\beta})$ of the estimated regression coefficients, the residual standard error $s(\hat{\varepsilon})$, and the residual autocorrelation ρ_1 are the same for the two regressions.

On the other hand, the key aspects of the regression results in tables 6 and 7 are similar for the two subperiods. The slope coefficients in the regressions of $S_{t+1} - S_t$ on $F_t - S_t$ are generally negative, which means that the coefficients in the complementary regressions of $F_t - S_{t+1}$ on $F_t - S_t$ are generally greater than 1.0. In the SUR tests, the hypothesis that all the slope coefficients in the $S_{t+1} - S_t$ regressions are 0.0 (or that the coefficients in the $F_t - S_{t+1}$ regressions are 1.0) is easily rejected in either subperiod.

Under the maintained hypothesis that the market assessments of E $(S_{t+1} - S_t)$ in $F_t - S_t$ are efficient or rational, the subperiod results confirm the earlier conclusions that (a) there is variation in both the P_t and E $(S_{t+1} - S_t)$ components of $F_t - S_t$, (b) the variance of the premium component of $F_t - S_t$ is large relative to the variance of the expected change in the spot rate, and (c) negative covariation between P_t and E $(S_{t+1} - S_t)$ dominates the variance of E $(S_{t+1} - S_t)$ to produce negative slope coefficients in the regressions of $S_{t+1} - S_t$ on $F_t - S_t$.

Interpretations

Various explanations of the results are suggested by the existing literature and by readers of earlier versions of this paper. Some of these explanations are discussed now. No explanation is necessarily complete, and they are not mutually exclusive. Moreover, generous readers of earlier drafts are not responsible for my paraphrasing of their comments.

An inefficient foreign exchange market

The interpretation of the results above is based on the hypothesis that the assessment of $E(S_{t+1} - S_t)$ in $F_t - S_t$ is efficient or rational. An alternative hypothesis is that the negative slope coefficients in the regressions of $S_{t+1} - S_t$ on $F_t - S_t$ reflect assessments of $E(S_{t+1} S_t)$ that are consistently perverse realtive to the true expected value of the change in the spot rate. The large positive coefficients in the $F_t - S_{t+1}$ regressions are then a simple consequence of the complementarity of the $F_t - S_{t+1}$ and $S_{t+1} - S_t$ regressions rather than manifestation of movement in rationally determined premiums. Under this interpretation, the similarity of the regression results for the two subperiods indicates that market irrationality in forecasting exchange rates is not cured by continued experience with flexible exchange rates.

Government intervention in the spot exchange market

A kind of 'market inefficiency', suggested by Richard Roll, can result from government intervention in the spot foreign exchange market. For example, suppose forward rates are determined by the interest rate parity condition (7) and interest rates in different countries rationally reflect their expected inflation rates. Left to the open market forces suggested by purchasing power parity, spot exchange rates would tend to move in the direction implied by the forward-spot differential $F_t - S_t$. Government logic and obstinacy, however, may be inversely related to natural market forces. Governments may support their currencies more vigorously (through open market operations, trade restrictions, and restrictions on capital flows) the stronger are the market forces, like differential expected inflation rates, which indicate that the currency should depreciate. They may try to move back toward a free market equilibrium by changing the direction of the underlying factors pressuring the exchange rate, like differential inflation rates, rather than by letting adjustments take place through the exchange rate.

The doomsday theory

Michael Mussa suggests that there are episodes, often brief, during which the distribution of anticipated changes in exchange rates is highly skewed. For example, market participants may assess a small probability that a country will change its monetary policy so that its inflation rate will rise dramatically relative to other countries. The result may be a highly skewed distribution of anticipated inflation rates, which in turn increases interest rate differentials and forward-spot exchange rate differentials between this country and other countries. Since the phenomenon centers on skewness that exists for brief periods, the *ex post* drawings from the distributions of anticipated inflation rates and changes in exchange rates are likely to be below the *ex ante* means. This creates negative sample correlations between changes in exchange rates and forward-spot differentials which would not be observed if the skewed distributions were sampled over longer periods.

Stochastic deviations from purchasing power parity

Stockman (1980) and Lucas (1982) develop international models in which shocks to real activity work in part through money demand functions to drive changes in inflation and exchange rates. Fama (1982) also argues that through the workings of a standard money demand function and inertia in money supply, variation in anticipated real activity in the U.S. leads to variation in expected inflation of the opposite sign. Fama and Gibbons (1982) argue that expected real returns on U.S. nominal bonds are also driven by and move in the same direction as anticipated real activity. With a somewhat different story in which monetary shocks cause changes in real variables, Tobin (1965) and Mundell (1963) likewise conclude that the expected real and expected inflation components of nominal interest rates are negatively correlated.

Suppose (a) interest rate parity holds; (b) expected changes in exchange rates reflect expected inflation differentials; and (c) the expected real components of nominal interest rates can vary somewhat independently across countries in response to purely domestic factors. These conditions, along with either the Tobin-Mundell or Fama-Gibbons stories for negative correlation between the expected real and expected inflation components of nominal interest rates, imply negative correlation between the premium, P_t, and the expected change in the spot rate, $E(S_{t+1} - = S_t)$, in the forward-spot differential, $F_t - S_t$.

To complete this story, however, we need a subplot to explain how the expected real returns on the nominal bonds of a country can vary in response to domestic factors that do not necessarily imply variation in the risks of the bonds. Segmented international capital markets can produce this result, but then the interest rate parity part of the story is likely to be lost. Alternatively, John Bilson suggests that such independent variation in the expected real returns on the nominal bonds of different countries can arise in open international capital markets when stochastic deviations from purchasing power parity (PPP) lead to strong preferences for borrowing and lending contracts denominated in one's domestic unit of account. Stulz (1981) provides a formal version of this kind of model in which deviations from PPP are due to the existence of nontraded goods. The Stulz model, in turn, can be viewed as a generalization of the Stockman (1980) and Lucas (1978, 1982) models.

Self-Assessment

1. **Choose the correct options:**

   (*i*) If the U.S. dollar depreciates in terms of the Euro:

      (*a*) European goods would be cheaper for Americans.

      (*b*) The relative price of U.S. exports would rise.

      (*c*) Americans would have to pay fewer dollars for one Euro.

      (*d*) American goods would be cheaper for Europeans.

   (*ii*) What accounts for most of the activity in the foreign exchange market?

      (*a*) Trading by financial institutions

      (*b*) Currency trade among central banks

      (*c*) Trading currency between importers and exporters

      (*d*) Interbank trading

   (*iii*) Which of the following is NOT a major currency trading center?

      (*a*) Frankfurt   (*b*) Tokyo

      (*c*) New York   (*d*) London

      (*e*) Chicago

   (*iv*) If a contract contains a promise that a specified amount of foreign currency will be delivered on the specified date in the future, this is:

      A swap.

      (*a*) A foreign exchange option.   (*b*) A futures contract.

      (*c*) A forward contract.   (*d*) A spot contract.

   (*v*) A saver will prefer asset X to asset Y if:

      (*a*) Asset X is less risky.   (*b*) Asset X is more liquid.

      (*c*) Asset X has a higher expected return. (*d*) All of the above.

      (*e*) None of the above.

14.3 Summary

- Large positive autocorrelations of the difference between the forward rate and the current spot rate indicate variation through time in either the premium component of $F_t - S_t$ or in the

assessment of the expected change in the spot rate. Moreover, slope coefficients in the regressions of $F_t - S_{t+1}$ and $S_{t+1} - S_t$ on $F_t - S_t$ that are reliably different from zero imply variation in both components of $F_t - S_t$. However, negative covariation between P_t and $E(S_{t+1} - S_t)$ leads to negative slope coefficients in the regressions of on $F_t - S_t$ and preempts accurate measurement of the variances of P_t and $E(S_{t+1} - S_t)$. Given market efficiency or rationality, the only conclusion we can draw from the negative slope coefficients in the $S_{t+1} - S_t$ regressions and slope coefficients greater than 1.0 in the complementary regressions of $F_t - S_{t+1}$ on $F_t - S_t$ is that the variance of the P_t component of $F_t - S_t$ is much larger than the variance of $E(S_{t+1} - S_t)$.

- Any forward rate can be interpreted as the sum of a premium and an expected future spot rate. Thus, our regression approach to examining the components of forward rates has broad applicability to financial and commodity market data. In Fama (1984), I apply the approach to forward and spot interest rates on U.S. Treasury bills, with somewhat more success. For example, unlike the forward exchange rate, which seems primarily to reflect variation in its premium component, the difference between the forward one month interest rate for one month ahead and the current one month spot interest rate, $F_t - R_t$, splits roughly equally between variation in its premium component and variation in the expected change in the one month spot interest rate. Moreover, in the interest rate data, $F_t - R_t$ sometimes has a larger variance than the *ex post* change in the one month spot interest rate, R_{t+1} R_t. Perhaps as a consequence, the *ex ante* $F_t - R_t$ explains from 15 to 70 percent of the variance of the *ex post* change in the spot interest rate, $R_{t+1} - R_t$. All of this is in striking contrast to the weak and somewhat perplexing picture that emerges from the exchange rate data, where variation in the *ex ante* forward-spot differential, $F_t - S_t$, is always small relative to the variation of the *ex post* change in the spot rate, $S_{t+1} - S_t$.

14.4 Key-Words

1. Exchange rate : Rate at which one may be converted into another. The exchange rate is used when simply converting one currency to another (such as for the purposes of travel to another country), or for engaging in speculation or trading in the foreign exchange market. There are a wide variety of factors which influence the exchange rate, such as interest rates, inflation, and the state of politics and the economy in each country, also called rate of exchange or foreign exchange rate or currency exchange rate.

14.5 Review Questions

1. What is the meaning of Exchange Rate? Explain.
2. Discuss the components exchange rate.

Answers: Self-Assessment

1. (*i*) (*a*)　　(*ii*) (*d*)　　(*iii*) (*e*)　　(*iv*) (*b*)　　(*v*) (*d*)

14.6 Further Readings

1. Bilson, John F.O., 1981, The speculative efficiency hypothesis. Journal of Business 54, July, 435–451.
2. Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, Journal of Financial Economics 7, Sept., 265–296.
3. Domowitz, Ian and Craig S. Hakkio, 1983, Conditional variance and the risk premium on the foreign exchange market, Manescript, Sept.
4. Fama, Eugene F. 1982, Inflation, output and money, Journal of Business 55, April, 201–231.

Unit 15 : Theories of Determination of Exchange Rate (PPP, Monetary)

```
CONTENTS
Objectives
Introduction
15.1  The Purchasing Power Parity Theory
15.2  Monetary Models of Exchange Rate Determination
15.3  Summary
15.4  Key-Words
15.5  Review Questions
15.6  Further Readings
```

Objectives

After reading this Unit students will be able to:

• Discuss the Purchasing Power Parity Theory.

• Explain the Monetary Models of Exchange Rate Determination.

Introduction

Many theories there have been written in respect to the main determinant of future exchange rates. Although the majority of these theories give adequate reasons in order to explain what actually determines the rates between the currencies, we can argue that there are many factors that may cause a currency fluctuation. Consequently, there is little that can be alleged in respect to the theory that better answers the question of what finally determines the exchange rates.

Here below, we will refer to the main theories regarding the determinants of the exchange rates.

15.1 The Purchasing Power Parity Theory

The purchasing power parity (PPP) theory was developed by Gustav Cassel in 1920 to determine the exchange rate between countries on inconvertible paper currencies. The theory states that equilibrium exchange rate between two inconvertible paper currencies is determined by the equality of their purchasing power. In other words, the rate of exchange between two countries is determined by their relative price levels. The theory can be explained with the help of an example.

Suppose India and England are on inconvertible paper standard and by spending Rs. 60, the same bundle of goods can be purchased in India as can be bought by spending £ 1 in England. Thus according to the purchasing power parity theory, the rate of exchange will be Rs. 60 = £ 1.

If the price levels in the two countries remain the same but the exchange rate moves to Rs. 50 = £ 1. This means that less rupees are required to buy the same bundle of goods in India as compared to £ 1 in England. It is a case of *overvaluation* of the exchange rate. This will encourage imports and discourage exports by India. As a result, the demand for pounds will increase and that of rupees will fall. This process will ultimately restore the normal exchange rate of Rs. 60 = £ 1. In the converse case, if the exchange rate moves to Rs. 70 = £ 1, the Indian currency become *undervalued*. As a result, exports are encouraged and imports are discouraged. The demand for rupees will rise and that for pounds will fall so that the normal exchange rate of Rs. 60 = £ 1 will be restored.

According to the theory, the exchange rate between two countries is determined at a point which expresses the equality between the respective purchasing powers of the two currencies. This is the purchasing power parity which is a moving par and not fixed par (as under the gold standard). Thus with every change in price level, the exchange rate also changes. To calculate the equilibrium exchange rate, the following formula is used :

$$R = \frac{\text{Domestic Price of a Foreign Currency} \times \text{Domestic Price Index}}{\text{Foreign Price Index}}$$

This is what the formula does. Let us explain it in terms of our above example. Before the change in the price level, the exchange rate was Rs. 60 = £ 1. Suppose the domestic (Indian) price index rises to 300 and the foreign (England) price index rises to 200, thus the new equilibrium exchange rate will be

$$R = \frac{£1 \times 300}{200} = £\ 1.5$$

or   Rs. 60 = £ 1.5

This will be the purchasing power parity between the two countries. In reality, the parity will be modified by the cost of transporting goods including duties, insurance, banking and other charges. These costs of transporting goods from one country to another are, in fact, the limits within the exchange rate can fluctuate depending upon the demand and supply of a country's currency. There is the upper limit, called the commodity export point; and the lower limit, known as the commodity import point. (These limits are not as definite as the gold points under the mint par theory).
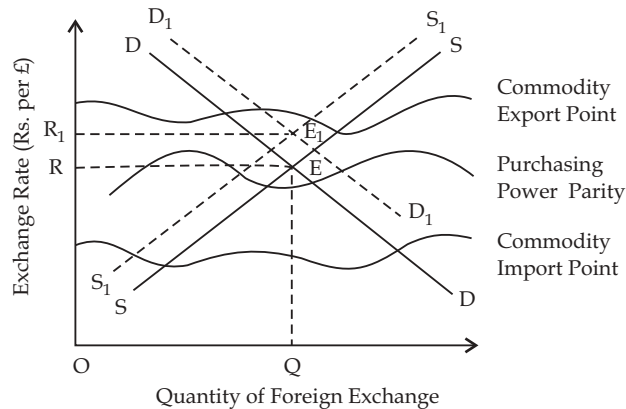

Figure 15.1

The purchasing power parity theory is illustrated in Figure 1 where DD is the demand curve for foreign currency (pound in our example) and SS is the supply curve of currency. OR is the rate to exchange of rupees per £ , which is determined by their intersection at point E so that the demand for the supply of foreign exchange equals OQ quantity. Suppose the price level rises in India and remains constant in England. This makes Indian exports costly in England and imports from England relatively cheaper in India. As a result, the demand for pounds increases and the supply of pounds decreases. Now the DD curve shifts upward to the right to D_1D_1 and the SS curve to the left to S_1S_1. The new equilibrium exchange rate is set at OR_1 rupees per pound, which represents the new purchasing power parity. The exchange rate rises by the same percentage as the India price level. The purchasing power curve shows that with relative change in the price levels, the exchange rate tends to fluctuate along this curve above or below the normal exchange rate. But there is a limit upto which the purchasing power parity curve can move up and down. The upper and lower limits are set by the commodity export point and the commodity import point respectively.

> *Did u know?* According to Cassel, the purchasing power parity is "determined by the quotients of the purchasing powers of the different currencies."

Its Criticisms. Cassell's PPP theory became very popular among economists during 1914-24 and was widely accepted as a realistic explanation of the determination of foreign exchange rate under inconvertible paper currencies. But it has been severely criticised for its weak theoretical base. Some of the criticisms are discussed as under :

1. One of the serious defects of the theory is that of calculating the price levels in the two countries. The use of index number in calculations presents many difficulties such as the base year, coverage and method of calculation. These may not be the same in both countries. The two countries may not include the same types of commodities in calculating the index numbers. Such difficulties make the index numbers only a rough guide for measuring the price levels and thus fail to give the correct purchasing power parity between the two countries.

2. According to the theory, the purchasing power parity between two countries is determined by comparing their general price levels. But the price level may be made up of internally traded plus internationally traded goods, or of the internationally traded goods. If the price level is calculated in terms of the internally traded goods, then the prices tend to equality in both countries, even allowing for the cost of transportation, tariffs, etc. Thus, according to Keynes, "confined to internationally traded commodities, the purchasing power parity becomes an empty truism." On the other hand, if the price level includes both internally and internationally traded goods, then price of internally traded goods may move in the opposite direction of internationally traded goods, at least in the short period. Thus the real exchange rate may not conform to the parities.

   Further, if the price level includes both types of goods, there is technical difficulty of people spending their money differently in the two countries, so that the basis for complete and accurate comparisons of price levels is lacking.

3. Another weakness of the purchasing power parity theory is that it applies to countries whose balance of payments is determined by the merchandise trade account. It is, therefore, not applicable to such countries whose exchange rate is influenced more by capital account.

4. The theory assumes the balance of payments to be in equilibrium in the base period for the determination of the new equilibrium exchange rate. This is a serious defect, because it is difficult to find the base year when the exchange rate was initially in equilibrium.

5. The theory is also based on the assumption that there have been no structural changes in the factors underlying the equilibrium in the base period. Such factors are changes in technology, resources, tastes, etc. This assumption is highly unrealistic because changes are bound to take place in these factors which, in turn, are likely to affect exchange rate.

6. The theory is based on the assumption of zero-capital movements. There are many items in the balance of payments such as insurance, shipping, and banking transactions, capital movements, etc. which are not affected by changes in the general price level. But these items affect the exchange rate by influencing the demand for and supply of foreign currencies. The theory is thus weak for it neglects the influence of these factors in determining the exchange rate.

7. The theory further assumes that changes in the price level bring about changes in exchange rates. But changes in exchange rates do affect the price level. For instance, if the external value of rupee falls, imports will become dearer. As a result, the costs and prices of goods using imported materials will rise. On the other hand, exports will become cheaper with fall in the external value of the rupee. Consequently, their demand will increase which will raise the demand for factors used for producing exports, and their prices will also rise. Thus changes in exchange rate do influence the price level.

8. Again, the theory assumes that the barter terms of trade do not change between the two trading

countries. This assumption is unrealistic because the barter terms of trade constantly change due to changes in the demand for foreign goods, in the volume of external loans, in the supply of exported goods, in transport Costs etc.

9. The theory is based on the assumption of free trade and laisser faire policy. But governments do not follow these policies these days. Rather, they impose a number of restrictions on the movement of goods between countries. Such trade restrictions are tariff, import quotas, customs duties and various exchange control devices which tend to reduce the volume of imports. These, in turn, cause wide deviations between the actual exchange rate and the exchange rate set by the purchasing Power Parity.

10. The equilibrium exchange rate may not be determined by the purchasing power parity between the two Countries. Rath, a sudden increase in the demand for goods of one country may raise the demand for its currency on the part of the other country. This will lead to a rise in the exchange rate.

11. Ragner Nurkse points out that the theory is one sided in that it is based exclusively on change in relative prices and neglects all factors that influence the demand for foreign exchange. The theory treats demand as a function of price but neglects the influence of aggregate income and expenditure on the volume and value of foreign trade, these are important factors which affect the exchange rate of a country.

Notes

According to keynes, one of the serious defects of this theory is that it fails to consider the elasticities of reciprocal demand. In fact, the exchange rate is determined not only by changes in relative prices, but also by the elasticities of recipocal demand between the two trading countries.

15.2 Monetary Models of Exchange Rate Determination

Since an exchange rate is the relative price of one nation's money in terms of the money of another nation, it is natural of think of an exchange rate as determined, at least proximately, by the outstanding stocks of these monies and by the demands to hold these stocks. This simple proposition is the starting-off point for two related but distinct classes of monetary models of exchange rate determination. The first class of monetary models, which have been widely applied in empirical studies of exchange rate behavior, expresses the current exchange rate as a function of the current stocks of domestic and foreign money and the current determinants of the demands for these monies, including domestic and foreign income and interest rates. The second class of monetary models, which has been more widely used in the-oretical work, focuses on the influence on the current exchange rate of the expected future path of money supplies and of factors affecting money demands. The distinguishing features of these two classes of models requires that they should be given separate attention.

The essential content of the first class of monetary models may be summarized in an equation of the form

$$e = m - m^* - (l[y, i, k] - l^* [y^*, i^*, k^*]), \tag{1}$$

where e is the logarithm of the price of foreign money in terms of domestic money, m is the logarithm of the domestic money supply, l is the logarithm of demand for domestic money (a function of domestic income, y, the domestic interest rate, i, and other factors k), and an asterisk (*) indicates variables for the foreign country. In some presentations, equation (15.1) is derived from the following assumptions: (1) The logarithm of the domestic price level, P, is determined by domestic money market equilibrium to be $P = m - l (y, i, k)$. (2) The logarithm of the foreign price level, P*, is determined by the foreign money market equilibrium condition to be $P^* = m^* - l^* (y^*, i^*, k^*)$. (3) The equilibrium exchange rate is determined by the requirement of purchasing power parity to be $e = P - P^* = m - m^* - (l [y, i, k] - l^* [y^*, i^*, k^*])$.

Monetary models of exchange rate determination have been criticized because of the inadequacy of the assumptions used to derive equation (1). In particular, the assumption of the purchasing power parity has been criticized as not consistent with the facts, especially the facts of the 1970s. The collapse of purchasing power parity in the 1970s, however, is not (in my judgment) adequate reason for rejecting equation (1) as a model (albeit an incomplete model) of exchange rate determination. This equation can be derived without explicit reference to purchasing power parity; indeed, it can be derived from a model that allows explicity for divergences from purchasing power parity. Moreover, some empirical studies employing equation (1) have noted that there are divergences from purchasing power parity and have argued that the conditions of money market equilibrium are more immediately relevant for determining the exchange rate (which is a freely adjusting asset price) than they are for determining national price levels. This, of course, leaves open the important question of what determines the behavior of national price levels, which in turn is an important element in explaining the behavior of real exchange rates. Nevertheless, if equation (1) worked well in explaining the behavior of nominal exchange rates, this form of monetary model of exchange rate determination would clearly make a sub-stantial contribution to our understanding of the economic forces influencing the behavior of exchange rates.

The principal empirical difficulty with this form of monetary model is that equation (1) does not work well in explaining actual movements in nominal exchange rates, unless we take into account shifts in the demands to hold different national monies that are difficult to explain in terms of traditional arguments appearing in money demand functions. An example illustrates this difficulty as well as a set of regressions. Between October 1976 and October 1980, the British pound appreciated by 50% in terms of the United States dollar, from \$1.60 to \$2.40. During this same period, monetary aggregates in Britain grew more rapidly than corresponding monetary aggregates in the United States, while real income (a key variable affecting the demand for money) grew less rapidly in Britain than in the United States. Of course, the increase in dollar value of sterling might be explained by an increase in the demand to hold sterling combined with a decrease in the demand to hold dollars, resulting from increased confidence in the future value of sterling (due to North Sea oil and the policies of Prime Minister Thatcher) and from increased concern about the inflationary consequences of the policies of the Carter administration. However, it is difficult to take these effects into account in a rigorous and disciplined fashion in an empirical version of equation (1).

Another important deficiency of equation (1) as a model of exchange rate determination is that it does not explicitly reveal the critical role of expectations of future economic conditions in determining the current exchange rate. From equation (1), there is no immediately apparent reason why changes in exchange rates should be largely random and unpredictable, or why new information that alters expectations about future economic conditions (including supplies and demands for national monies) should induce such random and unpredictable changes in exchange rates.

The second general class of monetary models of exchange rate determination does not suffer from this deficiency. These models usually treat a small or moderate size economy that takes conditions in the rest of the world as given. The critical condition determining the exchange rate for this country is the requirement of money market equilibrium;

$$m = k + \zeta \cdot e - \eta \cdot \mathrm{D}^{\mathrm{e}}(e), \; \zeta, \; \eta > 0, \qquad\qquad (2)$$

where m is the logarithm of the domestic money supply, e is the logarithm of the price of foreign money in terms of domestic money, k summarizes all exogenous factors affecting the logarithm of the demand for domestic money, and $\mathrm{D}^{\mathrm{e}}(e) = \mathrm{E}(e(t+1); t) - e(t)$ is the expected rate of change of the exchange rate. Equation (2) should be thought of as a reduced-form equilibrium condition derived from a more basic model of goods and asset market equilibrium. In this reduced form, the parameter ζ captures all of the mechanisms through which an increase in the price of foreign money increases the demand for domestic money, and the parameter η captures all of the mechanisms through which an increase in the expected rate of change of the price of foreign money affects the demand for domestic money.

Since the reduced-form demand for domestic money depends on the expected rate of change of the exchange rate, it follows that the current equilibrium exchange rate depends not only on the current values of m and k, but also on the expectation of next period's exchange rate;

$$e(t) = \left[1/(\zeta+\eta)\right]\cdot\left[m(t)-k(t)\right]+\left[\eta/(\zeta+\eta)\right]\cdot E(e(t+1);t). \qquad (3)$$

Forward iteration of (3), justified by the assumption of rational expectations, leads to the conclusion that the exchange rate expected at any future date is an exponentially weighted sum of expected future differences between m and k;

$$E(e(s);t) = \left[\zeta/(\zeta+\eta)\right]\cdot\sum_{j=0}^{\infty}\left[\eta/(\zeta+\eta)\right]^{j}\cdot E\big(w(s+j);t\big), \qquad (4)$$

were $w(u) = (1/\zeta)\cdot\left[m(u)-k(u)\right]$. The current exchange rate, $e(t) = E(e(t); t)$, is found by setting $s = t$ in (4). This result reveals the fundamental principle that the current exchange rate depends on the entire future expected path of differences between (the logarithms of) the money supply and the exogenous component of money demand.

Equation (4) may be used to decompose the change in the exchange rate into its expected and unexpected components. The expected change in the exchange rate is given by

$$D^{e}\left[e(t)\right] = (\zeta/(\zeta+\eta))\cdot\left[E(e(t+1);t)-E(w(t);t)\right] \qquad (5)$$

If, as is plausible to suppose, $\zeta/(\zeta+\eta)$ is on the order of one-tenth or one-twentieth, then large monthly expected changes in the exchange rate should be unlikely.[1] In contrast, the unexpected change in the exchange rate is given by

$$D^{u}\left[e(t)\right] = (\zeta/(\zeta+\eta))\cdot\sum_{j=0}^{\infty}(\eta/(\zeta+\eta))^{j}\cdot\left[E\big(w(t+j+l));t+1\big)\right. \qquad (6)$$

$$\left.- E\big(w(t+j+l);t\big)\right].$$

If the new information received between t and $t+1$ leads to a substantial revision of expectations concerning all future w's (in the same direction), this random and unpredictable component of the change in the exchange rate could be quite large.

To proceed with the analysis of changes in the exchange rate, it is necessary to specify how expectations about m and k are formed and revised. One convenient *theoretical* assumption is that k is a known constant, $\overline{k}$, that the money supply is observed each period before the exchange rate is determined, and that the stochastic process generating the money supply is known to economic agents and used by them (together with data on the present and past money supplies) to project the future course of the money supply. To be specific, suppose that m is generated by a random walk plus noise but that economic agents observe only m and not its permanent (random walk) and transitory (noise) components. In this case, economic agents will form an estimate $\hat{m}(t)$, of the current level of the permanent component of m by taking a weighted average of present and past m's, and they will attribute the difference, $m(t)-\hat{m}(t)$, to the present transitory component of m. The expected level of

---

1. In order to have an interest elasticity of money demand (given by $i\cdot\eta$) equal to 0.1, when the nominal interest rate is 1% per month, we must have $\eta = 10$. If $\zeta = 1$, as it would under strict purchasing power parity and no currency substitution, then $\zeta/(\zeta+\eta)$ would equal $1/11$. If the interest elasticity of money demand were as large as 0.2 and ζ were as small as 0.5, then $\zeta/(\zeta+\eta)$ would be as small as $1/41$.

m in any future period will equal $\hat{m}(t)$. The current exchange rate, $e\ (t) = (1/\zeta)\cdot\left[\hat{m}(t)-\bar{k}\right]+$ $\left[1/(\zeta+\eta)\right]\cdot\left[m(t)-\hat{m}(t)\right]$, fully reflects the component of the money supply that is thought to be permanent, but is less strongly affected by the component of the money supply that is thought to be transitory. The expected change in the exchange rate, $\mathbf{D}^{\mathbf{e}}\left[e(t)\right] = -\left[1/(\zeta+\eta)\right]\cdot\left[m(t)-\hat{m}(t)\right]$, reflects the expected disappearance of the transitory component of m. The information received by economic agents between t and $t+1$ is measured by difference between the actual level of $m\,(t+1)$ and the level that was expected at time t. E $(m\,(t+1);\ t) = \hat{m}(t)$. A fraction, α, of this difference is attributed to an increase in the premanent component of m, and the remaining fraction, $1-\alpha$, is attributed to the transitory component in $m\,(t+1)$, where the fraction α is an increasing function of the ratio of the variance of disturbances to the permanent component of m to the variance of transitory disturbances to the permanent component of m to the variance of transitory disturbances to m.

The unexpected change in the exchange rate,

$$\mathbf{D}^{\mathbf{u}}\left[e(t)\right] = \left\{(\alpha/\zeta)+\left[(1-\alpha)/(\zeta+\eta)\right]\right\}\ \cdot\left[m(t+1)-\hat{m}(t)\right],\ \text{reflects,}$$

as it should, the information received by economic agents between t and $t+1$. Consistent with common sense, this unexpected change in the exchange rate is greater the greater is then deviation of the money supply from its expected level and the greater is the fraction of this deviation that is attributed to a change in the permanent component of the money supply.

This example illustrates the key point that the nature of the stochastic process governing the behavior of the exchange rate depends on the process generating the behavior of the money supply and on the information about this process that is available to economic agents. In particular, this example illustrates that the response of the exchange rate to a change in the money supply depends on the extent to which this change was unanticipated and on the extent to which any unanticipated change is thought to indicate a permanent change in the money supply.

Aside from its theoretical usefulness, however, the assumption that economic agents use their knowledge of the (fixed) stochastic process generating the money supply as the primary ingredient in forming the expectations necessary for determining the exchange rate is not likely to provide a fully adequate empirical explanation of actual exchange rate movements. One likely reason for this inadequacy is that economic agents use many sources of information, other than the observed money supply series and other easily measured variables, in forming and revising their expectations concerning future money supply behavior. For example, the depreciation of the French franc on the day following the election of President Mitterand clearly was not due to any observed policy change (registered in the behavior of the money supply or other variables) since President Mitterand did not assume office until 3 weeks later. It must have been due to a change in expectations about future policy resulting from the fact of his election.

Another important barrier to monetary explanations of actual exchange rate movements arises from the lack of adequate measures of the exogenous factors affecting the demand for money and of expectations concerning the future behavior of these factors. Almost certainly, there have been shifts in the demands to hold national monies that are not accounted for either by changes in the traditional arguments appearing in money demand functions (such as levels of national income) or by changes in expectations about future exchange rate movements induced by changes in expectations about money supply behavior. In theory, such demand shifts should play a role of coordinate importance with changes in money supplies (and changes in expectations about future money) supplies in determining movements in exchange rages. The inadequacy of measures of money demand shifts means, therefore, that a substantial fraction of actual exchange rate movements will not be adequately explained by monetary models.

One possible way around this difficulty is to adopt the view that changes in exchange rates which cannot be explained by changes in the actual or expected behavior of money supplies must be due to

changes in the actual or expected behavior of money demands. The tautological view of the monetary model of exchange rate determination can be justified on the grounds that the money market equilibrium condition represented by equation (2) is a reduced form that incorporates all of the conditions of goods and asset market equilibrium. However, this tautological view of the monetary model still does not provide an explanation of many exchange rate movements, other than ascribing them to "shifts in money demands" arising from unknown sources. Moreover, while it is possible to view all economic forces affecting the exchange rate as operating through money demand or money supply, this may lead to a rather convoluted and unnatural view of the mechanisms through which some economic forces affect the exchange rate. In such circumstances, it is not sensible to insist on an exclusively monetary interpretation of the determination of exchange rates.

Self-Assessment

1. **Choose the correct options:**

    (*i*) The (uncovered) interest parity condition:

    (*a*) Takes into account the liquidity of the assets.

    (*b*) Involves forward exchange rate and spot exchange rate.

    (*c*) Describes the equilibrium in the foreign exchange market.

    (*d*) Takes into account the risk differential between the assets.

    (*ii*) If the interest rate on a deposit in Euros is 6% per year, and the Euro is expected to depreciate against the U.S. dollar by 1%, what does the interest parity condition imply about the interest rate on the deposit in U.S. dollars?

    (*a*) 7%

    (*b*) 5%

    (*c*) 6%

    (*d*) There is not enough information to find out.

    (*iii*) The interest parity condition involves four variables. Which one adjusts to ensure equilibrium?

    (*a*) Domestic interest rate

    (*b*) Foreign interest rate

    (*c*) Expected future exchange rate

    (*d*) Current exchange rate

    (*iv*) The U.S. dollar will appreciate if:

    (*a*) The Euro interest rate rises.

    (*b*) The U.S. interest rate falls.

    (*c*) The U.S. dollar is expected to depreciate.

    (*d*) The U.S. dollar is expected to appreciate

15.3 Summary

- The purchasing power parity (PPP) theory was developed by Gustav Cassel in 1920 to determine the exchange rate between countries on inconvertible paper currencies. The theory states that equilibrium exchange rate between two inconvertible paper currencies is determined by the equality of their purchasing power. In other words, the rate of exchange between two countries is determined by their relative price levels. The theory can be explained with the help of an example.

- The purchasing power parity theory is illustrated in Figure 1 where DD is the demand curve for foreign currency (pound in our example) and SS is the supply curve of currency. OR is the rate to exchange of rupees per £ , which is determined by their intersection at point E so that the demand for the supply of foreign exchange equals OQ quantity. Suppose the price level

rises in India and remains constant in England. This makes Indian exports costly in England and imports from England relatively cheaper in India. As a result, the demand for pounds increases and the supply of pounds decreases. Now the DD curve shifts upward to the right to D_1D_1 and the SS curve to the left to S_1S_1. The new equilibrium exchange rate is set at OR_1 rupees per pound, which represents the new purchasing power parity. The exchange rate rises by the same percentage as the India price level.

• Aside from its theoretical usefulness, however, the assumption that economic agents use their knowledge of the (fixed) stochastic process generating the money supply as the primary ingredient in forming the expectations necessary for determining the exchange rate is not likely to provide a fully adequate empirical explanation of actual exchange rate movements.

15.4 Key-Words

1. Depreciation : A noncash expense that reduces the value of an asset as a result of wear and tear, age, or obsolescence. Most assets lose their value over time (in other words, they depreciate), and must be replaced once the end of their useful life is reached. There are several accounting methods that are used in order to write off an asset's depreciation cost over the period of its useful life. Because it is a non-cash expense, depreciation lowers the company's reported earnings while increasing free cash flow.

2. Appreciation : An increase in the value of an asset over time. The increase can occur for a number of reasons including increased demand or weakening supply, or as a result of changes in inflation or interest rates. This is the opposite of depreciation, which is a decrease over time

15.5 Review Questions

1. What is the purchasing power parity theory? Discuss.
2. Determine the monetary models of exchange rate.

Answers: Self-Assessment

1. (*i*) (*c*)   (*ii*) (*b*)   (*iii*) (*d*)   (*iv*) (*d*)

15.6 Further Readings

1. Bilson, John F.O., 1981, The speculative efficiency hypothesis. Journal of Business 54, July, 435–451.

2. Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, Journal of Financial Economics 7, Sept., 265–296.

3. Domowitz, Ian and Craig S. Hakkio, 1983, Conditional variance and the risk premium on the foreign exchange market, Manescript, Sept.

4. Fama, Eugene F. 1982, Inflation, output and money, Journal of Business 55, April, 201–231.

Unit 16 : Theories of Determination of Exchange Rate (Portfolio and Balance of Payments)

```
CONTENTS
Objectives
Introduction
16.1  Foreign Exchange Regimes
16.2  Volatility and Risk
16.3  Determinants of Exchange Rate
16.4  Exchange Rate Forecasting
16.5  Summary
16.6  Key-Words
16.7  Review Questions
16.8  Further Readings
```

Objectives

After reading this Unit students will be able to:

- Discuss the Purchasing Power Parity Theory.
- Explain the Monetary Models of Exchange Rate Determination.

Introduction

By definition, the Foreign Exchange Market is a market in which different currencies can be exchanged at a specific rate called the foreign exchange rate. We can anticipate the huge importance of the foreign exchange rate if we can just consider the influence of it on the imports and exports of a country.

For example, let's assume a currency appreciation - the euro against the US dollar. Firstly, the exports of the European Union (E.U) nations will become 'expensive' for the United States of America (USA), which among other things means that E.U product will lose in terms of competitiveness. Secondly, such a currency appreciation will be to the benefit of E.U imports, should those be payable in US dollars. Conversely, a depreciation3 of the euro against the US dollar will cause an opposite impact.

On the other hand, the rapid growth of international trade (both the import penetration4 and the export ratio5) during the last decades, which was mainly due to the increase of the open economies, enhances the significance of the foreign exchange rates.

16.1 Foreign Exchange Regimes

Undoubtedly, governments have always paid very serious attention to the exchange rate of a country's currency, utilizing any available 'means' at hand, in order to stabilize the 'desirable' range of rate.

Historically, there were periods that governments through the central banks intervened in the foreign exchange market in order to affect the fluctuation of the exchange rate that otherwise would be determined by market forces. There were also periods with no intervention when the exchange rate, just like a price (Parkin M. and King D. 1992) was determined by supply and demand.

On 22nd July, 1944, at Bretton Woods in the United States of America, 44 countries agreed that a broad international action was necessary to maintain an international monetary system, which would promote foreign trade6. In this respect, it established a worldwide system of fixed exchange rates between currencies. Actually, the 'tool' was gold, with the following quota: one ounce of gold was to

be worth US dollars 35. After the establishment of the fixed rate of exchange, all other currencies were pegged to the US dollar at a fixed exchange rate.

As stated in the article7 'The End Of the "Fixed" Dollar', by the beginning of the 1960s, the US dollar 35 = 1 oz. gold ratio was becoming more and more difficult to sustain. Gold demand was rising and the U.S. Gold reserves were declining.

On 15th August, 1971, President Nixon, repudiated the international obligation of the U.S. to redeem its dollar in gold.

By the end of 1974, gold had soared from $35 to $195 an ounce.

Since the collapse of the Bretton Woods agreement (February 1973), the world's currencies have "floated" with respect to the US dollar.

Thus, the foreign exchange rate regime changed from a 'fixed exchange rate' to a 'flexible or floating exchange rate'. A system in which, exchange rates are determined by supply and demand that is called 'clean float' or where governments through central banks intervene (buy and sell currencies) in the markets, which is called 'dirty float'.

Notes   The demand for foreign exchange arises from the debit side of the balance of payments. it is equal to the value of payments made to the foreign country for goods and services purchased from it plus loans and investments made abroad.

16.2 Volatility and Risk

Undoubtedly, dramatic movements in the value of currencies can occur where the forces of supply and demand freely determine the price. Consequently, such a system increases the exchange rate risk associated with but not limited to international transactions.

The cross-border financial activity differs from the domestic activity in respect to related risk due to the fact that when investing in a foreign country you have to consider many other factors, such as:

* **Tax system:** differences related to the specific country's system.

* **Political risk:** a democratic country is preferable to a non-democratic one.

* **Government intervention**: it is also preferable to deal with a country without government intervention.

* **Business risk**: unforeseen changes in the general economic environment.

* In addition, the likely volatility in the exchange rate can drastically affect the cost, profits and return on investments of international firms, thus, resulting in the following levels of risk:

* **Economic exposure:** Transaction exposure is related to those activities that trade internationally. For example, a EU company imports bicycle components from the United States with 2 months' credit. Possible US dollar depreciation will be for the benefit of the EU organization because it will pay fewer euros. On the contrary, if the US dollar appreciates, the company will suffer a loss, due to the fact that it will pay more euros. The transaction exposure (the risk of adverse movements in the exchange rate) can be eliminated using hedging instruments. An example of this would be forward rate contracts.

* Operational exposure. Although a company may not trade globally, due to competitiveness, it may suffer the exchange rate risk. For example, a US bicycle producer will have a competitive advantage compared to a EU producer of a similar type of bicycle if the euro depreciates against the US dollar. The price of the US bicycle if converted into euros will fall, and consequently will attract EU members to buy it.

* Translation exposure Assuming a company that has a subsidiary outside the EU and expects profits in one year's time. Based on the current exchange rate between foreign and domestic

currency, the company has converted the amount of profits in its local currency. If the euro appreciates against the local currency, then the amount of profits when converted into euros will be less.

16.3 Determinants of Exchange Rate

Many theories there have been written in respect to the main determinant of future exchange rates. Although the majority of these theories give adequate reasons in order to explain what actually determines the rates between the currencies, we can argue that there are many factors that may cause a currency fluctuation. Consequently, there is little that can be alleged in respect to the theory that better answers the question of what finally determines the exchange rates.

Here below, we will refer to the main theories regarding the determinants of the exchange rates.

1. Supply and Demand

    As stated earlier, the exchange rate, just like commodities, determines its price responding to the forces of supply and demand8. Therefore, if for some reason people increase their demand9 (shift of the curve from D to D1) for a specific currency, then the price will rise from A to B, provided the supply remains stable. On the contrary, if the supply10 is increased (shift of the curve from S to S1), the price will decline from A to C, provided the demand remains stable .

    Any excess supply (above the equilibrium point) or excess demand (below the equilibrium point) will increase or decrease temporarily foreign currency reserves accordingly. Finally, such disequilibrium situations will be eliminated through the pricing, e.g. the market itself.

2. Purchasing Power Parity (PPP)

    By definition the PPP states that using a unit of a currency, let us say one euro, which is the purchasing power that can purchase the same goods worldwide. The theory is based on the 'law of one price', which argues that should a euro price of a good be multiplied by the exchange rate (€ /US$) then it will result in an equal price of the good in US dollars. In other words, if we assume that the exchange rate between the € and US $ states at 1/1.2, then goods that cost € 10 in the EU should cost US$ 12 in the United States. Otherwise, arbitrage11 profits will occur.

    However, it is finally the market that through supply and demand will force accordingly the euro and US dollar prices to the equilibrium point. Thus, the law of one price will be reinstated, as well as the purchase power parity between the euro and US dollar.

    Inflation differentials between countries will also be eliminated in terms of their effect on the prices of the goods because the PPP will adjust to equal the ratio of their price levels12. More specifically, as stated in their book (Lumby S. & Jones C. 1999) "the currency of the country with the higher rate of inflation will depreciate against the other country's currency by approximately the inflation deferential".

    In conclusion, it can be argued that the theory, although it describes in a sufficient way the determination of the exchange rates, is not of good value, mainly because of the following two disadvantages. Firstly, not all goods are traded internationally (for example, buildings) and secondly, the transportation cost should represent a small amount of the good's worth.

3. The Balance of Payments (BOP) Approach

    The balance of payments approach is another method that explains what the factors are that determine the supply and demand curves of a country's currency.

    As it is known from macroeconomics, the balance of payments is a method of recording all the international monetary transactions of a country during a specific period of time. The transactions recorded are divided into three categories: the current account transactions13, the capital account transactions14, and the central bank transactions15.

    The aforementioned categories can show a deficit or a surplus, but theoretically the overall payments (the BOP as a whole) should be zero - which rarely happens.

    As stated earlier, a currency's price depreciation or appreciation (the change in the value of money), directly affects the volume of a country's imports and exports and, consequently, a likely fluctuation in the exchange rates can add to BOP discrepancies.

For example, a likely depreciation will increase the value of exports in home currency terms (the larger the exports demand elasticity the greater the increase).

Conversely, the imports will become 'more expensive' and their value will be reduced in home currency (the larger the imports demand elasticity the greater the decrease).

Consequently, we can argue that unless the value of exports increases less than the value of imports, the depreciation will improve the current account. More specifically, we can finally assess the impact of the currency's depreciation on the current account only by considering the price sensitivity of imports and exports.

4. The Monetary Approach

    In this approach attention is given to the stock of currencies in comparison to the willingness of people to hold these stocks.

    According to the monetary theory, exchange rates adjust to ensure that the quantity of money in each currency supplied is equal to the quantity demanded (Parkin M. & King D. 1992).

    Both Quantity Theory of Money (QTM) and Purchasing Power Parity (PPP) have been used in support of the aforementioned theory.

    The QTM states that there is a direct relationship between the quantity of money and the level of prices of goods and services sold (Investopedia.com). In other words, more money equals more inflation.

    In a domestic framework, the following equation has been formulated16.

    MV = PY

    M: Money supply/demand

    V: Velocity of circulation (the number of times money change hands)

    P: Average price levels

    Y: GDP

    Finally, we can conclude that an increase in the money supply leads to inflation, which in turn results in the decrease in the value of money or purchasing power.

    Consequently, if we also consider this in an international context, we will appreciate the following implications:

    Firstly, a rapid increase in the money supply (in the home currency), which as stated earlier means inflation, will put into effect the PPP resulting in the depreciation of the currency's exchange rate.

    Secondly, a higher interest rate will also result in the currency's depreciation because of the positive relationship between interest rates and money circulation.

    Finally, if the GDP grows faster than overseas GDP, the demand for money will increase. Assuming there is a given supply of money, the exchanged rate will decrease, which is in direct contrast to the PPP approach.

5. The Portfolio Balance Approach

    The portfolio balance approach takes into consideration the diversification of investors' portfolio assets. Diversification is a technique that attempts to reduce risk by investing both among various financial instruments and across national borders, to mention just a few.

    For example, here below we consider a combination of domestic and foreign money and domestic and foreign bonds.

Both the M and the B lines show combinations of domestic interest rates and exchange rates. The upward line M is in agreement with the equilibrium in the money market and the downward line B is in agreement with the equilibrium in the bond market. Point E, which is the intersection of M with B, represents the combination of interest rate with the exchange rate that gives equilibrium to both the money and bond markets.

What the theory argues is that an increase in the money supply will lead to a depreciation of the exchange rate. The extent of the depreciation depends upon the slope of the curves M and B.

For example, if we consider an increase in the domestic money supply, we will anticipate that a lower interest rate and / or a higher exchange rate can only absorb the excess supply, which in turn will result in the reduction of bonds. To this end, line M will move to the right and line B will move to the left.

16.4 Exchange Rate Forecasting

In the previous section we referred to various theories in respect to the main determinant of future exchange rates. However, it is actually an empirical topic to identify the most important factors. As a result, we can argue that forecasting exchange rates is a difficult task because of so many factors that might be regarded as determinants. In addition, using formulas17 (similar to the Interest Rate Parity Theorem) will result in only imprecise estimates of future currencies exchange, basically because the data related to inflation differentials and future spot rates are by themselves only estimates.

On the other hand, it remains doubtful weather or not the foreign exchange market is an efficient18 one - since it seems that not all information but only the historical is incorporated in prices (weak form efficiency).

Although we share the opinion that there is no reliable method available to forecast exchange rates, we will refer here below to some of the main concepts.

1. The Unbiased Expectations Hypothesis

    The concept of the Unbiased Expectations Hypothesis argues that the forward rate is an unbiased forecaster of the future spot rate. Thus, the forward rate at time t for maturity at time T must equal the markets expectation at time t for maturity at time T.

    FtT = Et(ST)

    In addition, the forward price is the expected spot price minus a risk premium to cover likely interest rate differentials.

    FtT = Et(ST) - ?. ? is the risk premium

2. Unbiased or Biased Predictor

    Finally, the UEH does not give a quite satisfactory reply as to whether or not a forward rate may be a biased predictor of a future spot rate.

    In this respect, the following equation gives adequate answer to the aforementioned question.

    ST = bo + b1 Ft,T + b2 I

    St: represents the realized spot rate for the maturity date

    FT: represents the forward rate

    I: represents any available information that affects the exchange rate

    b2: is a statistically significant that represents the market's efficiency if all the information is not incorporated in the forward rate (inefficient market)

    Studying historical data, we can conclude that in the majority of instances the bo does not equal to zero and the b1 does not equal to 1, which means that the forward rate is actually a biased predictor of the future spot rate. Contrarily, both the bo and the b1 have a negative value, demonstrating that there is a risk premium (mentioned earlier) related to the forward rate.

3. Purchasing Power Parity

    In accordance with the PPP concept, the inflation differentials between countries affect the exchange rate, and consequently the PPP could be of good value in order to forecast the exchange rate. In Cochran & Defina study (Cochran & Defina 1995), they show that the exchange rates, although they deviate from the PPP, they finally return to their PPP levels. However, during their deviation the possibility of the exchange rates moving backward from or forwards to the PPP remains the same, and consequently the PPP did not prove itself to be a consistent or reliable 'tool' for forecasting.

4. Econometric Models

   The econometric models in their attempt to forecast the exchange rates take into consideration other factors that are regarded as determinants of the exchange rates. More specifically: inflation, the relative19 GDP levels, the relative interest rates, and the relative money supply.

   Although the econometric models perform well in general, we can argue that these models do not lead to reliable forecasts because they use in many cases 'past' data.

   In addition, it can be argued that empirical studies indicate that the ability of such models to forecast is greater in the long term than in the short term.

5. Technical Analysis Forecasting

   In contrast to the fundamental analysis that considers financial data, the technical analysis analyses diagrams based on past data. More specifically, it accepts the following:

   The market is efficient (all information is incorporated into the prices).

   The prices are moved by trend.

   The history repeats itself.

The technical analysis takes into consideration indicators, such as moving averages and oscillators, the volume of transactions, trend lines, and time or price filters, and it combines their trend according to some charting rules.20Then in turn the technical analysis tries to identify the future price and anticipate the price patterns in the financial and commodity markets.

The applied chartist techniques, either on bar charts or on candlestick21 charts, lead to more reliable forecasts when dealing with short-term periods.

Self-Assessment

1. Choose the correct options:

   (*i*) If the U.S. dollar depreciates in terms of the Euro:

   (*a*) American goods would be cheaper for Europeans.

   (*b*) Americans would have to pay fewer dollars for one Euro.

   (*c*) The relative price of U.S. exports would rise.

   (*d*) European goods would be cheaper for Americans.

   (*ii*) What accounts for most of the activity in the foreign exchange market?

   (*a*) Interbank trading

   (*b*) Currency trade among central banks

   (*c*) Trading currency between importers and exporters

   (*d*) Trading by financial institutions

   (*iii*) Which of the following is NOT a major currency trading center?

   (*a*) Chicago   (*b*) Tokyo

   (*c*) London   (*d*) New York

   (*e*) Frankfurt

   (*iv*) What is the "arbitrage" opportunity in the foreign exchange market?

   (*a*) A cross-rate.

   (*b*) A difference between the exchange rate for buying and selling the currency from the same bank.

   (*c*) A difference between the exchange rates in different trading centers.

   (*d*) A fee that brokers charge for trading currency of their clients.

   (*v*) If a contract contains a promise that a specified amount of foreign currency will be delivered on the specified date in the future, this is:

(*a*) A forward contract.   (*b*) A swap.

(*c*) A futures contract.   (*d*) A foreign exchange option.

(*e*) A spot contract.

(*vi*) A saver will prefer asset X to asset Y if:

(*a*) Asset X is more liquid.   (*b*) Asset X has a higher expected return.

(*c*) Asset X is less risky.   (*d*) All of the above.

(*e*) None of the above.

16.5 Summary

• In this Unit we initially analyzed the theories of exchange rate determination, and we concluded that the exchange rate of a currency, just like commodities, determine its price responding to the forces of supply and demand.

• However, it remains difficult to evaluate the 'weight' of each factor that influences the supply and demand flows. In this respect, we can argue that in the long term fundamental forces drive the currency's movements, e.g., inflation rates, interest rates, and GDP levels, to mention just a few, while in the short run, news and events drive the exchange rates movements. Moreover, we still remain unsure as to the ranking of the factors that finally determine the exchange rates.

• We have further considered various concepts and models thought to be reliable forecasters of the exchange rates. We have concluded that the forward rates are not good predictors of the future spot rates. In other words, the forward rate does not provide an 'unbiased' estimate of the exchange rate movements.

• Taking into consideration the time horizons, we come to the conclusion that fundamental economic forces, such as purchasing power parity and the balance of payments, did not automatically affect the exchange rates, but they require a considerable amount of time. As a result they are not reliable over short time periods. Conversely, we can argue that the technical analysis performs comparatively well in the short run since news and events are incorporated into the diagrams and trends.

• We finally conclude the following: Firstly, the foreign exchange market is not an efficient one. Otherwise, it would be futile to beat or try to forecast the market. Secondly, because of so many determinants of the exchange rates, it is difficult to proceed to a reliable estimation for future rates. Thirdly, there is no unfailing method available to forecast exchange rates, and we have further determined that the forward rate provides a biased estimate of the future spot rate. As a result, companies trading internationally face exposure to exchange rate risk.

• In this respect, the mentioned 'costly' hedging22 instruments in Appendix 1 have been developed in order to manage the exposure related to unfavorable currency movements. More specifically, it can be argued that there is no 'perfect' hedge or in other words a hedge with 100 % efficiency. What in practice happens is that a small profit or loss is made. For example, when an importing company has a liability that is payable in six months, it can hedge its exposure to exchange rate risk by using a forward contract. e.g., the company can buy the due amount at the six months forward rate. The company having used the aforementioned hedging instrument has no 'uncertainty' about the exchange rate movements because it will pay a specific and previously known amount in its own currency. In other words, the company has 'locked' its exposure, and consequently, its cost irrespective of whether or not the after six months 'spot' rate proved to be in favour of or not in favour of his decision to hedge.

• Finally, we wish to underline that as indicated earlier, there are no unfailing methods to forecast exchange rates; consequently, the companies face an exposure related to the currencies' fluctuation. However, their exposure to exchange rates movements can be managed, controlled, and even more, eliminated by using hedging instruments.

16.6 Key-Words

1. Portfolio approach : An approach to explaining exchange rates that stresses their role in changing the proportions of different currency-denominated assets.

   A grouping of financial assets such as stocks, bonds and cash equivalents, as well as their mutual, exchange-traded and closed-fund counterparts. Portfolios are held directly by investors and/or managed by financial professionals.

16.7 Review Questions

1. What do you mean by portfolio? Discuss portfolio approach.

2. Discuss the theory to determine exchange rate .

3. Write a short note on the balance of payment theory.

Answers: Self-Assessment

1. (*i*) (*d*)   (*ii*) (*a*)   (*iii*) (*a*)

    (*iv*) (*c*)   (*v*) (*c*)   (*vi*) (*e*)

16.8 Further Readings

Books

1. Dornbusch, Rudiger and Stanley Fischer (1980). \Exchange Rates and the Current Account." American Economic Review 70, 960 {71.

2. Frenkel, Jacob A. and Michael L. Mussa (1985). \Asset Markets, Exchange Rates, and the Balance of Payments: The Reformulation of Doctrine." See Jones and Kenen (1985), pp. 679{747.

3. Isaac, Alan G. (1989). \Wealth E_ects and the Current Account with Endogenous Terms of Trade." Journal of Macroeconomics 11(4), xxx.

4. Jones, Ronald W. and Peter B. Kenen, eds. (1985). Handbook of International Economics, Volume 2. Amsterdam: North Holland Publishing Co.

5. Kenen, Peter (1985). \Macroeconomic Theory and Policy: How the Closed Economy was Opened." See Jones and Kenen (1985), Chapter 13, pp. 625{677.

6. Kouri, Pentti J.K. (1976, May). \The Exchange Rate and the Balance of Payments in the Short Run and the Long Run: A Monetary Approach." Scandinavian Journal of Economics 78(2), 280{304.

Unit 17 : Process of Adjustments : Gold Standard, Fixed Exchange Rates and Flexible Exchange Rate

---

CONTENTS

Objectives

Introduction

17.1 The Gold Standard

17.2 Fixed Exchange Rates

17.3 Flexible Exchange Rates

17.4 Summary

17.5 Key-Words

17.6 Review Questions

17.7 Further Readings

Objectives

After reading this Unit students will be able to:

• Know the Gold Standard.

• Discuss the Fixed Exchange Rate and Flexible Exchange Rate.

Introduction

This process reestablishes equilibrium in the foreign exchange market. The Bank of England loses foreign reserves since it is forced to buy pounds and sell gold to keep the pound price of gold fixed. Foreign central banks gain reserves as they *buy* gold with their currencies. Countries share equally in the burden of balance of payments adjustment. Because official foreign reserves are declining in Britain and increasing abroad, the British money supply is falling, pushing the British interest rate back up, and foreign money supplies are rising, pushing foreign interest rates down. Once interest rates have again become equal across countries, asset markets are in equilibrium and there is no further tendency for the Bank of England to lose gold or for foreign central banks to gain it. The total world money supply (not the British money supply) ends up being higher by the amount of the Bank of England's domestic asset purchase. Interest rates are lower throughout the world.

Our example illustrates the symmetric nature of international monetary adjustment under a gold standard. Whenever a country is losing reserves and seeing its money supply shrink as a consequence, foreign countries are gaining reserves and seeing their money supplies expand. In contrast, monetary adjustment under a reserve currency standard is highly asymmetric. Countries can gain or lose reserves without inducing any change in the money supply of the reserve currency country, and only the latter country has the ability to influence domestic and world monetary conditions.[1]

---

1. Originally, gold coins were a substantial part of the currency supply in gold standard countries. A country's gold losses to foreigners therefore did not have to take the form of a fall in central bank gold holdings : Private citizens could melt gold coins into ingots and ship them abroad, where they were either reminted as foreign gold coins or sold to the foreign central bank for paper currency. In terms of our earlier analysis of the central bank balance sheet, circulating gold coins are considered to make up a component of the monetary base that is not a central bank liability. Either form of gold export would thus result in a fall in the domestic money supply and an increase in foreign money supplies.

17.1 The Gold Standard

An international gold standard avoids the asymmetry inherent in a reserve currency standard by avoiding the "Nth currency" problem. Under a gold standard, each country fixes the price of its currency in terms of gold by standing ready to trade domestic currency for gold whenever necessary to defend the official price. Because there are N currencies and N prices of gold in terms of those currencies, no single country occupies a privileged position within the system : Each is responsible for pegging its currency's price in terms of the official international reserve asset, gold.

The Mechanics of a Gold Standard

Because countries tie their currencies to gold under a gold standard, official international reserves take the form of gold. Gold standard rules also require each country to allow unhindered imports and exports of gold across its borders. Under these arrangements, a gold standard, like a reserve currency system, results in fixed exchange rates between all currencies. For example, if the dollar price of gold is pegged at $35 per ounce by the Federal Reserve while the pound price of gold is pegged at £14.58 per ounce by Britain's central bank, the Bank of England, the dollar/pound exchange rate must be constant at ($35 per ounce) ÷ (£14.58 per ounce) = $2.40 per pound. The same arbitrage process that holds cross exchange rates fixed under a reserve currency system keeps exchange rates fixed under a gold standard as well.

Symmetric Monetary Adjustment Under a Gold Standard

Because of the inherent symmetry of a gold standard, no country in the system occupies a privileged position by being relieved of the commitment to intervene. By considering the international effects of a purchase of domestic assets by one central bank, we can see in more detail how monetary policy works under a gold standard.

Suppose the Bank of England decides to increase its money supply through a purchase of domestic assets. The initial increase in Britain's money supply will put downward pressure on British interest rates and make foreign currency assets more attractive than British assets. Holders of pound deposits will attempt to sell them for foreign deposits, but no *private* buyers will come forward. Under floating exchange rates, the pound would depreciate against foreign currencies until interest parity had been reestablished. This depreciation cannot occur when all currencies are tied to gold, however. What happens ? Because central banks are obliged to trade their currencies for gold at fixed rates, unhappy holders of pounds can sell these to the Bank of England for gold, sell the gold to other central banks for their currencies, and use these currencies to purchase deposits that offer interest rates higher than the interest rate on pounds. Britain therefore experiences a private financial outflow and foreign countries experience an inflow.

> **Notes** A gold standard therefore places automatic limits on the extent to which central banks can cause increases in national price levels through expansionary monetary policies.

Benefits and Drawbacks of the Gold Standard

Advocates of the gold standard argue that it has another desirable property besides symmetry. Because central banks throughout the world are obliged to fix the money price of gold, they cannot allow their money supplies to grow more rapidly than real money demand, since such rapid monetary growth eventually raises the money prices of all goods and services, including gold. These limits can make the real values of national monies more stable and predictable, thereby enhancing the transaction economies arising from the use of money. No such limits to money creation exist under a reserve currency system; the reserve currency country faces no automatic barrier to unlimited money creation.

Offsetting this potential benefit of a gold standard are some drawbacks :

1. The gold standard places undesirable constraints on the use of monetary policy to fight unemployment. In a worldwide recession, it might be desirable for all countries to expand their money supplies jointly even if this were to raise the price of gold in terms of national currencies.

2. Tying currency values to gold ensures a stable overall price level only if the *relative* price of gold and other goods and services is stable. For example, suppose the dollar price of gold is $35 per ounce while the price of gold in terms of a typical output basket is one-third of a basket per ounce. This implies a price level of $105 per output basket. Now suppose that there is a major gold discovery in South America and the relative price of gold in terms of output falls to one-fourth of a basket per ounce. With the dollar price of gold unchanged at $35 per ounce, the price level would have to rise from $105 to $140 per basket. In fact, studies of the gold standard era do reveal surprisingly large price level fluctuations arising from such changes in gold's relative price.

3. An international payments system based on gold is problematic because central banks cannot increase their holdings of international reserves as their economies grow unless there are continual new gold discoveries. Every central bank would need to hold some gold reserves to fix its currency's gold price and serve as a buffer against unforeseen economic mishaps. Central banks might thereby bring about world unemployment as they attempted to compete for reserves by selling domestic assets and thus shrinking their money supplies.

4. The gold standard could give countries with potentially large gold production, such as Russia and South Africa, considerable ability to influence macroeconomic conditions throughout the world through market sales of gold.

   Because of these drawbacks, few economists favor a return to the gold standard today. While most central banks continue to hold some gold as part of their international reserves, the price of gold now plays no special role in influencing countries' monetary policies.

Did u know?   As early as 1923, the British economist John Maynard Keynes characterized gold as a "barbarous relic" of an earlier international monetary system.

The Bimetallic Standard

Up until the early 1870s, many countries adhered to a **bimetallic standard** in which the currency was based on both silver and gold. The United States was bimetallic from 1837 until the Civil War, although the major bimetallic power of the day was France, which abandoned bimetallism for gold in 1873.

In a bimetallic system, a country's mint will coin specified amounts of gold *or* silver into the national currency unit (typically for a fee). In the United States before the Civil War, for example, 371.25 grains of silver (a grain being l/480th of an ounce) or 23.22 grains of gold could be turned into a silver or, respectively, gold dollar. That mint parity made gold worth 371.25/23.22 = 16 times as much as silver.

The mint parity could differ from the market relative price of the two metals, however, and when it did, one or the other might go out of circulation. For example, if the price of gold in terms of silver were to rise to 20 : 1, a depreciation of silver relative to the mint parity of 16 : 1, no one would want to turn gold into gold dollar coins at the mint. More dollars could be obtained by instead using the gold to buy silver in the market, and then having the silver coined into dollars. As a result, gold would tend to go out of monetary circulation when its relative market price rose above the mint relative price, and silver coin would tend to disappear in the opposite case.

The advantage of bimetallism was that it might reduce the price level instability resulting from use of one of the metals alone. Were gold to become scarce and expensive, cheaper and relatively abundant silver would become the predominant form of money, thereby mitigating the deflation that a pure gold standard would imply. Notwithstanding this advantage, by the late nineteenth century most of the world had followed Britain, the leading industrial power of the day, onto a pure gold standard.

The Gold Exchange Standard

Halfway between the gold standard and a pure reserve currency standard is the **gold exchange standard.** Under a gold exchange standard central banks' reserves consist of gold *and* currencies whose prices in terms of gold are fixed, and each central bank fixes its exchange rate to a currency with a fixed gold price. A gold exchange standard can operate like a gold standard in restraining excessive monetary growth throughout the world, but it allows more flexibility in the growth of international reserves, which can consist of assets besides gold. A gold exchange standard is, however, subject to the other limitations of a gold standard listed above.

The post-World War II reserve currency system centered on the dollar was, in fact, originally set up as a gold exchange standard. While foreign central banks did the job of pegging exchange rates, the U.S. Federal Reserve was responsible for holding the dollar price of gold at $35 an ounce. By the mid-1960s, the system operated in practice more like a pure reserve currency system than a gold standard. For reasons explained in the next chapter, President Nixon unilaterally severed the dollar's link to gold in August 1971, shortly before the system of fixed dollar exchange rates was abandoned.

The Demand for International Reserves

The unit explained that a central bank's assets are divided between domestic-currency assets, such as domestic government bonds, and foreign-currency assets, the bank's international reserves. Historically and up to the present day, international reserves have been prized by central banks because they can be traded to foreigners for goods and services even in circumstances, such as financial crises and wars, when the value of domestic assets may come into doubt. Gold played the role of international reserve asset *par excellence* under the gold standard—and economists debate whether the United States dollar plays that role today and, if so, for how long that unique American privilege can last. Because central banks and governments may alter their policies to affect national holdings of international reserves, it is important to understand the factors that influence countries' demands for international reserves.

A good starting point for thinking about international reserves is the model in the chapter in which domestic and foreign bonds are perfect substitutes, the exchange rate is fixed, and confidence in the fixed exchange rate is absolute. In that model, our result that monetary policy is ineffective also implies that individual central banks can painlessly acquire all the international reserves they need! They do so simply by an open-market sale of domestic assets, which immediately causes an equal inflow of foreign assets but no change in the home interest rate or in other domestic economic conditions. In real life matters may not be so easy, because the circumstances in which countries need reserves are precisely those in which the above conditions of perfect confidence in creditworthiness and in the exchange-rate parity are likely to be violated. As a result, central banks manage their reserves in a *precautionary* manner, holding a stock they believe will be sufficient in future times of crisis.

As usual there are costs as well as benefits of acquiring and holding reserves, and the level of reserves that a central bank wishes to hold will reflect a balance between the two. Some monetary authorities (such as that of Hong Kong) value reserves so highly that the entire money supply is backed by foreign assets—there are no domestic monetary assets at all. In most cases, however, central banks hold both domestic and foreign assets, with the optimal level of reserves determined by the tradeoff between costs and benefits.

Starting in the mid-1960s, economists developed and sought empirical verification of formal theories of the demand for international reserves. In that setting, with international capital markets much more limited than they are today, a major threat to reserves was a sudden drop in export earnings, and central banks measured reserve levels in terms of the number of months of import needs those reserves could cover. Accordingly, the variability levels of exports, imports, and international financial flows, all of which could cause reserves to fluctuate too close to zero, were viewed as prime determinants of the demand for international reserves. In this theory, higher variability would raise the demand for reserves. An additional variable raising the average demand for reserves might be the adjustment cost countries would suffer if they suddenly had to reduce exports or raise imports to

generate a trade suplus, or raise interest rates to draw in foreign capital. Higher economic openness could make such adjustments easier, thereby reducing the demand for reserves, but might also make an economy more vulnerable to foreign trade shocks, thereby raising desired reserve holdings.

On the other hand, the main cost of holding reserves is their interest cost. A central bank that switches from domestic bonds to foreign reserves loses the interest on the domestic bonds and instead earns the interest on dollars. If markets harbor any fears that the domestic currency could be revalued, then domestic bonds will offer a higher interest rate than foreign reserves, implying that it is costly to switch the central bank's portfolio toward reserves. In addition, reserves may offer lower interest simply because of their higher liquidity.

It was argued in the 1960s that countries with more flexible exchange rates would find it easier to generate an export surplus if reserves ran low—they could allow their currencies to depreciate, perhaps avoiding the recession that might otherwise be needed to create a trade balance surplus. When industrial countries moved to floating exchange rates in the early 1970s, many economists therefore expected that the demand for international reserves would drop sharply.

Figure 1 shows, however, that nothing of the sort happened. For industrial countries, the growth rate of international reserves has declined only slightly since the 1960s. Industrial-country reserves have persistently grown at roughly the same pace as nominal industrial-country income. For developing countries, the growth rate of reserves has, if anything, risen (though the recent sharp upsurge is to some degree a reflection of huge reserve purchases by China). Accelerating reserve growth has taken place despite the adoption of more flexible exchange rates by many developing countries.
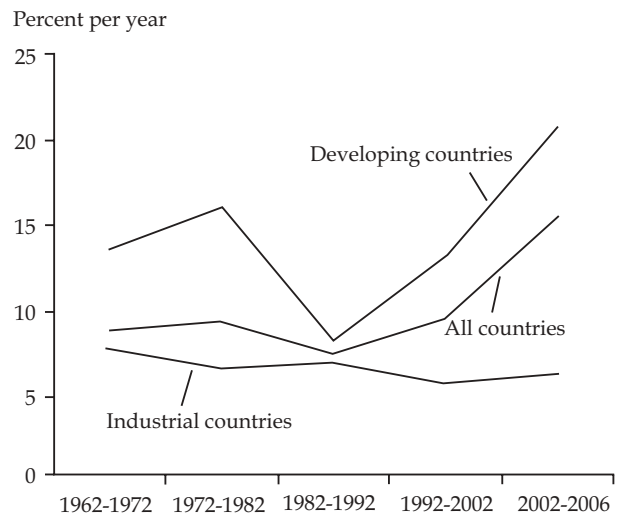

Figure 17.1: Growth Rates of International Reserves

Annualized growth rates of international reserves did not decline sharply after the early 1970s. Recently, developing countries have added large sums to their reserve holdings.

Source : Economic Report of the President, 2007.

One explanation for this development, which we will discuss further in later chapters, is that the growth of global capital markets has vastly increased the potential variability of financial flows across national borders, and especially across the borders of crisis prone developing countries. The sharp decline in developing-country reserve growth in the 1982-1992 period, shown in the figure, reflects an international debt crisis during the years 1982-1989. In that crisis, foreign lending sources dried up and developing countries were forced to draw on their reserves. The episode illustrates well why developing countries have added so eagerly to their reserve holdings. Even a developing country with a floating exchange rate might need to pay off foreign creditors and domestic residents with dollars to avoid a financial crisis and a currency collapse.

Nothing about this explanation contradicts earlier theories. The demand for international reserves still reflects the variability in the balance of payments. The rapid globalization of financial markets in recent years has, however, caused a big increase in potential variability and in the potential risks that variability poses.

17.2 Fixed Exchange Rates

In a system of fixed exchange rates, the exchange rate is fixed at an official predetermined rate. The central bank acts as a market maker and steps in to fill any imbalance between demand and supply.

Such a system has two major advantages. The first is that fixed exchange rates imply reduced uncertainty, and this helps reduce the costs of international trade transactions. The second is that fixed exchange rates act as to discipline monetary authorities, preventing them from pursuing inflationary policies. This argument was emphasized when Brazil and Argentina adopted fixed-exchange-rate-based arrangements in the 1990s. The logic is that excessive money supply expansion generates inflation that, in turn, gives agents an incentive to shift into currencies with purchasing power that is not being eroded. Such shifts force the central bank to intervene and buy the currency to protect the exchange rate, thereby reducing the money supply. In this fashion, fixed exchange rates establish an automatic mechanism that prevents central banks from excessive money supply expansion, and central banks are forced to tighten the money supply whenever inflation starts to increase to levels that will spur currency flight. This mechanism is referred to as a nominal anchor, with the exchange rate serving to anchor the price level. In countries with histories of excessive inflation and where central banks have lost credibility with financial markets, it is argued that employing a fixed-exchange-rate nominal anchor is a good way to win back credibility. Moreover, the costs of such commitment are small if monetary policy is viewed as being unable to systematically impact equilibrium real interest rates and the equilibrium level of real economic activity.

Balanced against these advantages are several disadvantages. First, by committing to a fixed exchange rate, a country gives up having the exchange rate as a shock absorber that helps insulate against external economic shocks. Second, the fixed exchange rate limits the ability to use domestic monetary policy to stabilize the economy, but, as noted above, this loss can be beneficial in cases where monetary authorities have a credibility problem due to past high inflation.

Third, the nature of the adjustment process under fixed exchange rates may have a significant deflationary bias. Abstracting from capital flows, countries with trade surpluses will experience an excess demand for their currencies, while countries with trade deficits will experience an excess supply of their currencies. If a deficit country is forced to keep buying its currency to defend the exchange rate, this leads to domestic monetary contraction in the deficit country, while the money supply of the surplus country increases due to the selling of foreign reserves by the deficit country. The classical macroeconomic assumption is that reductions in the money supply cause prices to decline but have no impact on output. This is the "neutrality of money proposition," whereby output and employment are determined by real economic factors (tastes, resources, and productive technology) and not by the amount of circulating paper (i.e., money). Applied to the global economy with fixed exchange rates, these money supply changes cause prices to fall in the deficit country and rise in the surplus country, thereby altering relative competitiveness and eliminating the trade deficit. However, such global monetarist reasoning is contested by Keynesian analysis that argues that monetary contraction induces real output contraction that is worsened by price deflation due to debt effects. The net result is that the adjustment process under fixed exchange rates causes domestic output contraction that ricochets back into the international economy, as falling domestic income causes reduced imports, in turn, reducing aggregate demand and income in other countries.

One possible way to avoid this contractionary outcome is to require the surplus country to defend its currency and prevent it from appreciating, rather than require the deficit country to do the defending. In this case, the system is prone to an expansionary bias, because the surplus country increases its money supply to prevent appreciation. However, this arrangement removes the discipline of fixed exchange rates on central banks. A second option for reducing contractionary bias is to have periodic discrete adjustments of the fixed exchange rate to eliminate fundamental trade imbalances. This was the Bretton Woods approach. However, it also removes (or at least significantly weakens) the discipline of fixed exchange rates on central banks. Additionally, it does away with the certainty of fixed exchange rates and invites market speculation aimed at anticipating or forcing a devaluation.

One claimed advantage of fixed exchange rates is that they reduce price uncertainty, which is good for international trade. However, introducing international capital mobility into a system of fixed exchange rates dramatically changes this conclusion. As noted earlier, capital mobility introduces portfolio and wealth allocation concerns that impact currency markets. Most importantly, capital mobility introduces financial market behaviors of speculation and herding into currency markets. These behaviors can render a fixed exchange-rate system financially fragile. If a country has a persistent

trade deficit, the central bank will be obliged to intervene to defend the exchange rate. Given finite holdings of foreign reserves, market participants will recognize that the central bank will eventually run out of foreign reserves with which to conduct this defense. At this stage, market participants may start selling to get out before the central bank runs out of reserves and is forced to devalue. As a result, the collapse can be brought forward in time, even when a central bank still has large reserve holdings. More importantly, speculators may begin to speculate against any currency they believe "subjectively" to be weak. In modern financial markets, speculators can raise enormous amounts of leverage that dwarf the foreign reserves of the central bank. Therefore, they can engage in a war of attrition that they can win as long as the central bank with weak currency is the one forced to defend the exchange rate. In effect, fixed exchange rates offer speculators a form of "one-way" option. If they speculate and win, they reap the huge reward of devaluation : if the central bank fights off the speculative attack, all they have lost are the transactions costs and interest for a short period, and these transaction costs are increasingly small due to technological innovations in electronic commerce. The upshot is that fixed exchange rates are fragile in a world with international capital mobility. This means that there is always a risk of speculatively induced collapse, and reducing that risk requires that countries hold large quantities of costly foreign reserves.

A final problem with fixed exchange rates concerns their impact on private-sector borrowing decisions, particularly in developing countries. Fixed exchange rates create a moral hazard, whereby agents think there is no currency risk associated with foreign currency borrowing. Agents, therefore, over-borrow foreign currency, and sudden collapses of the exchange rate can leave them saddled with huge debt burdens measured in domestic currency terms. At this stage, a country can be plunged into a cycle of debt deflation and economic contraction, as happened in East Asia and Argentina.

17.3 Flexible Exchange Rates

In a flexible system, the exchange rate is determined by market forces of demand and supply for a currency. Among economists, there is a generic presumption that markets are stable, and that the actions of agents as represented by demand and supply are based on rational decisions predicated on "economic fundamentals," and that market economies (i.e., the full network of individual markets that make up the economy) have a propensity to adjust smoothly and rapidly to full employment equilibrium in the absence of market impediments (i.e., inappropriate regulations and restrictions on price adjustment). This generic presumption predisposes economists to look favorably on flexible exchange rates.

The principle advantage of flexible exchange rates concerns their ability to insulate and stabilize economic activity. With regard to external shocks, the exchange rate can adjust to maintain trade balance. Thus, if export demand declines, the exchange rate can depreciate to lower export prices and restore demand. In effect, the external sector can be balanced by adjusting one price (the exchange rate) rather than adjusting thousands of prices, which would be necessary if restoring balance through downward aggregate price and nominal wage adjustment. In addition, a flexible exchange rate can help in the adjustment to internal demand shocks. Thus, a domestic boom will tend to raise domestic interest rates, thereby attracting financial inflows and driving up the exchange rate. This appreciation will tend to reduce export demand and switch consumption away from domestically produced goods to imports, thereby reducing aggregate demand and cooling the boom.

A second major advantage of flexible exchange rates is that they strengthen the power of monetary policy, which can be used to ensure domestic economic balance. Thus, in recession, the monetary authority can lower interest rates, thereby causing financial capital to exit, which depreciates the exchange rate and stimulates net exports.

Balanced against these advantages are some disadvantages. First, flexible exchange rates imply exchange-rate uncertainty that raises the cost of international trade to the extent that firms hedge this uncertainty. The greater the volatility of exchange rates, the greater the uncertainty and cost. Perhaps even more important is that exchange-rate uncertainty may cause firms to diversify sources of production internationally to protect against exchange-rate changes that can adversely affect their costs and competitive positions. This hedge-driven diversification is inefficient, being driven by

uncertainty rather than production efficiency concerns. Moreover, firms may end up with overall excess capacity that they are willing to carry as a hedge against exchange-rate exposure. This raises costs. Additionally, to the extent that internationally diversified production increases firms' bargaining power with labor, the distribution of income may be tilted away from wages to profits. Volatile flexible exchange rates may, thereby, have contributed to the adverse income distribution outcomes associated with globalization.

A second problem with flexible exchange rates relates to the issue of capital mobility. In the absence of capital mobility, demand and supply in exchange markets will reflect the balance of trade. Countries running surpluses will experience excess demand for their currency, as their trading partners seek to obtain currency to pay for imports, and this will cause the surplus country currency to appreciate. Conversely, currencies of deficit countries will tend to depreciate, as they sell their currency to get surplus country currency. This is the double-entry logic of market exchange. Every purchase is matched by an offer of exchange. In currency markets, the match is one currency for another. If the Marshall-Lerner elasticity conditions are met, the depreciation of the deficit country's exchange rate will tend, over time (after J-curve effects have worked through), to restore trade balance, which will then cause its currency to stop depreciating. Under such conditions, the foreign exchange market is stable.

However, given capital mobility, demand and supply in exchange markets will reflect more than just trade balance considerations. They will also reflect asset portfolio considerations and decisions to hold wealth across different national financial markets. This brings an asset market dimension to foreign exchange markets that can be highly problematic. In particular, currency markets will take on the character of asset markets. As such, they may be volatile and subject to speculative manias and herd behaviors. This opens the way for asset market volatility to impact exchange rates and, thereby, impact output and employment. Thus, as financial investors move money into a country, they will appreciate the exchange rate. This can make industries uncompetitive, resulting in plant closures and job losses despite the absence of any change in factory floor productivity. Capital inflows will also drive up asset prices and lower interest rates, thereby promoting asset-centered booms and distorting the allocation of resources.

In the event that the inflows reverse, the result can be a collapse in asset prices and a rise in interest rates, as happened in East Asia in 1997. Flexible exchange rates plus unrestricted capital mobility can, therefore, make a volatile cocktail.

Self-Assessment

1. **Choose the correct options:**

   (*i*) The British economist John Maynard Keynes characterized gold as a barbarous relif in ....

   (*a*) 1901   (*b*) 1923

   (*c*) 1920   (*d*) None of these

   (*ii*) The United States was bimetallic from 1837 until the Civil War, although the major bimetallic power of the day was ...., which abandoned bimetallisus for gold in 1873.

   (*a*) Germany   (*b*) France

   (*c*) Itly   (*d*) None of these

   (*iii*) President Nixon unilaterally severed the dollar's link to gold in August, ...

   (*a*) 1971   (*b*) 1965

   (*c*) 1980   (*d*) 1951

   (*iv*) Brazil and Argentina adopted fixed-rate-based arrangements in ....

   (*a*) 1970s   (*b*) 1990s

   (*c*) 1920s   (*d*) None of these.

17.4 Summary

- This process reestablishes equilibrium in the foreign exchange market. The Bank of England loses foreign reserves since it is forced to buy pounds and sell gold to keep the pound price of gold fixed. Foreign central banks gain reserves as they *buy* gold with their currencies.

- Our example illustrates the symmetric nature of international monetary adjustment under a gold standard. Whenever a country is losing reserves and seeing its money supply shrink as a consequence, foreign countries are gaining reserves and seeing their money supplies expand. In contrast, monetary adjustment under a reserve currency standard is highly asymmetric.

- A gold standard therefore places automatic limits on the extent to which central banks can cause increases in national price levels through expansionary monetary policies. These limits can make the real values of national monies more stable and predictable, thereby enhancing the transaction economies arising from the use of money. No such limits to money creation exist under a reserve currency system; the reserve currency country faces no automatic barrier to unlimited money creation.

- In a bimetallic system, a country's mint will coin specified amounts of gold *or* silver into the national currency unit (typically for a fee).

- The mint parity could differ from the market relative price of the two metals, however, and when it did, one or the other might go out of circulation. For example, if the price of gold in terms of silver were to rise to 20 : 1, a depreciation of silver relative to the mint parity of 16 : 1, no one would want to turn gold into gold dollar coins at the mint. More dollars could be obtained by instead using the gold to buy silver in the market, and then having the silver coined into dollars.

- The unit explained that a central bank's assets are divided between domestic-currency assets, such as domestic government bonds, and foreign-currency assets, the bank's international reserves. Historically and up to the present day, international reserves have been prized by central banks because they can be traded to foreigners for goods and services even in circumstances, such as financial crises and wars, when the value of domestic assets may come into doubt.

- In real life matters may not be so easy, because the circumstances in which countries need reserves are precisely those in which the above conditions of perfect confidence in creditworthiness and in the exchange-rate parity are likely to be violated. As a result, central banks manage their reserves in a *precautionary* manner, holding a stock they believe will be sufficient in future times of crisis.

- Some monetary authorities (such as that of Hong Kong) value reserves so highly that the entire money supply is backed by foreign assets—there are no domestic monetary assets at all. In most cases, however, central banks hold both domestic and foreign assets, with the optimal level of reserves determined by the tradeoff between costs and benefits.

- In a system of fixed exchange rates, the exchange rate is fixed at an official predetermined rate. The central bank acts as a market maker and steps in to fill any imbalance between demand and supply. Such a system has two major advantages.

- The second is that fixed exchange rates act as to discipline monetary authorities, preventing them from pursuing inflationary policies. This argument was emphasized when Brazil and Argentina adopted fixed-exchange-rate-based arrangements in the 1990s. The logic is that excessive money supply expansion generates inflation that, in turn, gives agents an incentive to shift into currencies with purchasing power that is not being eroded. Such shifts force the central bank to intervene and buy the currency to protect the exchange rate, thereby reducing the money supply. In this fashion, fixed exchange rates establish an automatic mechanism that prevents central banks from excessive money supply expansion, and central banks are forced to tighten the money supply whenever inflation starts to increase to levels that will spur currency flight.

- Third, the nature of the adjustment process under fixed exchange rates may have a significant deflationary bias. Abstracting from capital flows, countries with trade surpluses will experience an excess demand for their currencies, while countries with trade deficits will experience an excess supply of their currencies. If a deficit country is forced to keep buying its currency to defend the exchange rate, this leads to domestic monetary contraction in the deficit country, while the money supply of the surplus country increases due to the selling of foreign reserves by the deficit country.

- A final problem with fixed exchange rates concerns their impact on private-sector borrowing decisions, particularly in developing countries. Fixed exchange rates create a moral hazard, whereby agents think there is no currency risk associated with foreign currency borrowing. Agents, therefore, over-borrow foreign currency, and sudden collapses of the exchange rate can leave them saddled with huge debt burdens measured in domestic currency terms.

17.5 Key-Words

1. Equilibrium : A state in which oppositing forces or influences are balanced.
2. International reserves : Foreign exchange reserves in a strict sense are only the foreign currency deposits and bonds.

17.6 Review Questions

1. What is the mechanics of a gold standard? Explain.
2. What are the benefits and drawbacks of the gold standard? Explain.
3. What are the demand for International Reserves? Discuss.

Answers: Self-Assessment

1.　(*i*)　(*b*)　　　　(*ii*)　(*b*)　　　　(*iii*)　(*a*)　　　　(*iv*)　(*b*)

17.7 Further Readings

Books

1. Dornbusch, Rudiger and Stanley Fischer (1980). \Exchange Rates and the Current Account." American Economic Review 70, 960 {71.

2. Frenkel, Jacob A. and Michael L. Mussa (1985). \Asset Markets, Exchange Rates, and the Balance of Payments: The Reformulation of Doctrine." See Jones and Kenen (1985), pp. 679{747.

3. Isaac, Alan G. (1989). \Wealth E_ects and the Current Account with Endogenous Terms of Trade." Journal of Macroeconomics 11(4), xxx.

4. Jones, Ronald W. and Peter B. Kenen, eds. (1985). Handbook of International Economics, Volume 2. Amsterdam: North Holland Publishing Co.

5. Kenen, Peter (1985). \Macroeconomic Theory and Policy: How the Closed Economy was Opened." See Jones and Kenen (1985), Chapter 13, pp. 625{677.

6. Kouri, Pentti J.K. (1976, May). \The Exchange Rate and the Balance of Payments in the Short Run and the Long Run: A Monetary Approach." Scandinavian Journal of Economics 78(2), 280{304.