The handle http://hdl.handle.net/1887/20601 holds various files of this Leiden University dissertation.

Author: Kinuthia, Bethuel Kinyanjui  
Title: Reversed fortunes in the south : a comparison of the role of FDI in industrial development in Kenya and Malaysia  
Issue Date: 2013-03-07
The role of foreign direct investment in economic development: A survey of the literature

Introduction
This chapter reviews the literature on the role of FDI in economic development. FDI has registered remarkable growth over the past several decades, exceeding the growth of world production and the growth of international trade. In addition, FDI in recent times has typically accounted for more than 60 per cent of private capital flows to developing countries (Herzer et al. 2008). The developments in the world economy have generated new international production networks and complex systems, which have resulted in new forms of FDI, necessitating a more integrated approach in the study of FDI among various disciplines. These developments have led to many publications focusing on the why, when, how, where, and causes of FDI occurrence, and its effects on economic growth and development. These studies can be broadly considered as dealing with the determinants of FDI and the effects of FDI on economic growth and development. An attempt is made in this chapter to review this literature and examine the policy implications. Attention is focused on the main currents of thought rather than giving coverage to all the publications.

In theory, as observed in the introductory chapter, there are several potential ways in which FDI can promote economic growth and development. First, FDI can finance capital formation through increases in capital stock. This is especially important in developing countries where savings do not match investments, creating a savings-investment gap. Second, FDI is generally assumed to be more
productive than domestic investment since it encourages the incorporation of new technologies in the production function of the host economy. By inviting MNEs to invest within their national boundaries, host countries hope to gain access to technologies and skills they do not yet possess. Since technology is viewed to some extent as a public good, FDI operations in host countries can result in technology transfer and spillovers.

These FDI-related technology spillovers are supposed to offset the effects of diminishing returns to capital and keep the economy on a long-term growth path. Moreover, endogenous growth models imply that FDI can promote long-term growth by augmenting the existing capital stock of knowledge in the host economy, through labour training and skill acquisition on the one hand and through the introduction of alternative management practices and organization arrangements on the other (de Mello 1997; Blomström & Kokko 1998). FDI is also considered more stable that other forms of capital flows such as Official Development Assistance (ODA), remittances, and portfolio capital flows. With this realization, most developing countries are desperate to remove restrictions on FDI inflows, while international donors and development agencies have focused ever more on promoting private rather than public flows as a catalyst for economic development.

Since the 1980s, scholars have integrated economic growth theories with the theory of trade, resulting from an interest in the study of the interactions between developed countries (North) and the developing countries (South). Krugman (1981) among the early pioneers showed that inequalities between the North and the South were due to the lack of industrialization of the latter. Others have attributed divergence in growth not only to increased product varieties or increasing product quality, as given by the endogenous growth models, but also to changes in the world trade patterns over the years (Romer 1990; Grossman & Helpman 1991).

Following the emergence of newly industrialized countries in Asia, it is now recognized that developing countries have increasingly specialized in exports of new products like the North, and in some countries such as Malaysia they comprise over half of their manufactured exports. This has resulted in the development of a new breed of trade models, which emphasize increasing returns to scale, imperfect competition, and the production of different varieties of new products to explain this new pattern. These models have complemented the traditional trade models, whether Hecksher-Ohlin or Ricardian comparative models (Chui et al. 2002), in explaining why intra-industry trade occurs in parallel to inter-industry trade. Owing to the realization that MNEs account for a significant proportion of trade, these studies have been extended to study FDI activity (ibid.).
The development within the industrial organization theories of the firm have also contributed to a better understanding of the FDI activity. These theories, while making a distinction between FDI and portfolio flows, suggest that FDI exists as a result of market imperfections and possesses unique advantages, which allow MNEs to compete with domestic firms in the host countries. These market failure approaches have led to the formalization and further developments in the form of transaction costs, internalization, and eclectic paradigms. These aspects have been incorporated in the Dunning (OLI) eclectic paradigm for analyzing FDI activity, where MNEs possess ownership (O), location (L), and internalization (I) advantages, which enable them to make investment decisions in host countries (Dunning 2001). All or some aspects of the eclectic paradigm have been incorporated in the trade and growth theories, giving rise to new trade theories, which have generated important insights as scholars seek to understand what causes FDI to go to one country and not the other, in addition to analyzing the North-South interactions (Helpman 1984; Markusen 1984; Dollar 1986; Chui et al. 2001).

While most of the theories developed within the neoclassical framework have emphasized potential benefits of FDI and trade in the growth and welfare of host countries, the proponents of uneven development have argued quite the opposite. In their view, international trade and FDI activity between the North and South is inherently unequal. The structuralists attribute this imbalance to the existence of structural rigidities, lags, and other characteristics contained in southern economies, which affect economic adjustment and the choice of development policies. As a result, international specialization along the lines of comparative advantage excludes the South from technical progress owing to their specialization in the export of primary commodities, while technology is concentrated in industry. In addition, the relative prices of primary commodities in terms of manufacturing prices are on a decline, further hurting the South (Prebisch 1950; Singer 1950).

Furthermore, countries in the South are considered trapped in poverty owing to low capital accumulation requiring huge investments to facilitate takeoff (Krugman 1981). However, in recent times there is a renewed interest in the neoclassical thinking by the new institutionalists, who have argued that host countries can benefit from FDI – if the state is involved in improvement of their domestic absorption capacities and puts in place institutions that can provide incentives leading to the generation of various spillovers from MNEs (Lall 2004; Lall & Narula 2004; Rodrik 2004).

---

1 The international capitalistic system, through both foreign trade and investment, systematically favours certain types of countries over others, as well as certain types of workers. This has the effect of widening the gap between the North and the South.
At the extreme end are the dependency theorists who see FDI as a form of neo-colonialism, where the North exploits the South. They blame the international capitalistic system, particularly its economic aspects, for countries in the South remaining poor. In addition, this literature tends to put more weight on the interaction of political and economic forces, in explaining underdevelopment in the South (dos Santos 1970). New developments in the dependency theories have been extended to the issues of social welfare and the environment. In this literature, MNEs can have damaging effects on the social welfare by exploiting labour and polluting the environment in the South (Oetzel & Doh 2009). Due to the potentially negative effects of FDI on host countries, dependency theories argue that FDI does not develop the South.

Since there are many competing theories explaining various aspects of FDI, this chapter will classify them into five groups, as presented in Table 2.1. The first is a mixture of the old growth and old trade models. The second group is models of FDI as presented by the industrial organization. The third group incorporates the elements in the first and two groups to generate old growth and “new” trade theories, which incorporate features of the industrial organization such as increasing returns to scale and product innovation. In the first and the third group, the processes driving growth are exogenous to the model. Fourth is a combination of new trade theories and new (endogenous) growth theories. Finally, in the last group are the theories of uneven development, which discuss the effects of FDI on social welfare and the natural environment.

Table 2.1 Theories explaining FDI activity

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Industrial Organization (Business School): The Dunning (OLI) eclectic paradigm (Dunning 2000)</td>
</tr>
<tr>
<td>3</td>
<td>Old Growth, New Trade: Dollar (1986)</td>
</tr>
<tr>
<td>5</td>
<td>Uneven Development theories: Hobson (1902), Lenin (1916)</td>
</tr>
</tbody>
</table>

The rest of the chapter is organized as follows: In each section, a brief description of a core model that exemplifies the type is set out, and a general survey of other papers in the same category is provided. The core old growth and old trade models are taken from Brems (1970), Findlay (1978), and Burgastaller & Saaverdra-Rivano (1984) and are discussed in the next section. This is followed by an exposition of the industrial organization theories of FDI as presented in the
Dunning (OLI) eclectic paradigm (Dunning 2000). The old growth, new trade model of Krugman (1979), generalized by Dollar (1986), is then discussed, followed by the new growth, new trade model as exemplified by Grossman & Helpman (1991), generalized by Chui, Levine & Pearlman (2001). The summary and conclusion section in this chapter is preceded by a presentation of the theories of uneven development.

Old growth and old trade models

One of the earliest growth-trade models focusing on FDI was presented by Brems (1970), which built and solved a two-country in the North, four-sector model of FDI. Unlike the earlier trade theories where FDI was considered as part of international trade capital, FDI in this model was defined as the movement of a bundle of money capital plus technological and managerial knowledge from a parent firm to its foreign subsidiary. The model relaxed the factor price equalization assumption, which underpinned the Heckscher-Ohlin (H-O) model of the neoclassical trade theory. The H-O model was based on a 2 x 2 x 2 general equilibrium framework with two countries (home and foreign), two factors of production (usually capital and labour), and two goods, assuming perfectly competitive goods and factor markets, identical constant returns to scale production functions, zero transport costs, and factor endowments that are such as to exclude specialization. According to the H-O model, commodities differed in relative factor intensities and countries differed in relative factor endowments, leading to international factor price differentials. Hence, a relatively capital-abundant country at home would either export the capital extensive good to the foreign country or, in the absence of commodity trade, move capital to the foreign country where returns on capital are higher and returns on labour lower, until factor price equalization is achieved.

Relaxing the twin assumption of commonly known production functions and non-specialization, factor price equalization was not expected. Consequently, incentives for factor movements existed. Brems assumed that labour could not move internationally at all, and only money capital could flow. His view was based on the US experience, where money capital was known to flow more easily and in larger volumes than in any other form of portfolio. Each of the two countries in the model contained two sectors, domestic – and foreign-owned. Entrepreneurs in each of the countries produced a unique good, using the neoclassical production function of the Solow (1956) type, where technology was assumed to be exogenous. Based on this production function this good was produced first, in the foreign country, by labour employed by the home country entrepreneurs, and second, by the immortal capital stock of that good. An act of investment by the home country would then comprise entrepreneurs of the home
country setting aside part of their output for installation there and this would comprise FDI. Entrepreneurs were believed to allocate their investment between the parent firm and foreign subsidiary in such a way as to maximize the present worth of all their future profits. Brems argued that if entrepreneurs at home were disposed towards highly automated methods, they would be so disposed in their foreign subsidiaries as well. Thus, technology progress made at home would be applied in the foreign subsidiaries.

In addition, the four commodities produced were assumed to be goods but not perfect substitutes in consumption, and each country’s consumption had a taste for all the four goods. The money wage rates were assumed to be the same in both sectors in each country, but the prices of the commodities varied, and as a consequence, the real wage rate was also varied in terms of any of the four goods. Furthermore, the exchange rate was also considered variable, and this reliance on the price mechanism allowed Brems to specify equilibrium conditions. He concluded by observing that capital movement in the two-way international direct investment is a permanent, realistic, and non-paradoxical feature of the full equilibrium. One country’s entrepreneurs may simply be superior in computers, while in another country they may be superior in steel. Once the two-way FDI flows are determined, so is their net effect in these countries. The direction of the net effect depended on the entire structure of the model. With complete symmetry, there was zero net effect of national capital export; but by breaking this assumption – assuming superiority in one country due to either thriftiness, technology, or consumer preferences for products offered by a country’s customers – the net effect of national capital export changes from zero to positive, suggesting that each of these causes of asymmetry may contribute to an explanation of US investment abroad.

Similarly, Findlay (1978) constructed a simple dynamic model that captured some aspects of the way in which the transfer of technology took place, including the role of FDI. The model consisted of two regions, a relatively backward and an advanced region, and employed the idea usually associated with Veblen (1915) and Gerschenkron (1962) that the greater the backlog of available opportunities to exploit – measured by the distance between the advanced and the backward region’s current levels of development – the greater the pressure for change within the backward region and the faster its rate of growth. In addition, diffusion of technology was considered as an analogy with the spread of a contagious disease, an idea initially proposed by Arrow (1971). The basis of the analogy is the fact that technical innovations are most effectively copied when there is personal contact between those who already have knowledge of the innovation and those who eventually adopt it. These considerations led to the hypothesis that, other things being equal, the rate of change of technical efficiency in the
backward region is an increasing function of the relative extent to which activities of the firm with their superior technology pervade the economy.\(^2\)

Findlay assumed a production function where foreign and domestic capital were essentially distinct factors of production, each with its own separate rate of returns, with no necessity for equalization. In addition, the rate of technological change in the backward region was expressed as a function of the degree to which it is exposed to foreign capital. This was unlike in MacDougall (1960), where the output of a country was determined by a production function with two inputs – labour and a homogenous stock of capital – and investment by foreigners was conceived of as a marginal additional stock of capital, which has the effect of increasing the domestic wage rate and reducing the domestic returns of capital, and there was no possibility of technology transfer.

The domestic economy in the relatively backward country was characterized by ‘surplus labour’, as described in the Lewis (1954) dual economy model.\(^3\) The foreign sector within this economy has superior technology with a higher rate of profit than the domestic sector. Moreover, it is assumed that the foreign sector pays a higher wage than the domestic sector. Based on this model, technological progress benefits this economy by increasing efficiency in production in both sectors at the early stages of development. This leads to domestic capital accumulation, since it is assumed to be partly financed by a tax on foreign profits. As the country progresses, capital accumulation in the domestic sector becomes sufficiently high, and this in turn raises the domestic rate of technological progress. This continued increase in relative domestic efficiency is, however, not favourable for foreign presence in the economy and hence diminishes as the economy approaches the steady state. Thus, Findlay argued that it is the necessity for self-reliance that is more conducive to rapid technological borrowing and that the presence of foreign capital retards rather than accelerates the process.

Findlay (1980), using similar concepts of advanced (North) and backward (South) regions, developed a trade-growth model in which the terms of trade emerged as the mechanism linking growth rates of output in the North and the South. The structural difference in the determination of the growth rate in the two regions produced asymmetrical consequences on the terms of trade of changes in technology and propensity to save. All improvements in the production function (Solow type) of the North increased its propensity to save, and the terms of trade remained unchanged in the long run and increased its real per-capita income. In

\(^2\) The rate of penetration was measured as the ratio of capital stock of foreign-owned (and managed) firms in the backward economy to the capital stock of the domestic-owned firms.

\(^3\) Lewis postulated a dual sectoral economy structure between agriculture and manufacturing in a closed economy setting. The agricultural sector is characterized by surplus labour; workers are paid an exogenously fixed real wage rate equal to the average product of labour, and all wages are consumed. These notions go back to earlier ideas of a wage fund theory in the classical economics of the early nineteenth century.
the South, on the other hand, these shifts led to a proportionate fall in the long-term terms of trade and a decline in real per-capita income measured in terms of manufacturers. These unfavourable effects, however, were compensated by an increase in the relative employment ratios if the elasticity of import demand is sufficiently high. As we shall see in the last section, this asymmetry is a prominent feature in the writings of Prebisch & Singer.

Burgastaller & Saaverdra-Rivano (1984) extended Findlay’s framework by introducing a link between the North and the South through perfect capital mobility of real capital and corresponding continuous equality of profits across regions, in addition to goods flow. In their model the size of the labour force in the North is fixed and fully employed in the short term. Southern entrepreneurs, on the contrary, face an infinitely elastic labour supply that keeps the real wage constant at subsistence level. Employment in the South accordingly is a variable, and will depend on the fraction of the accumulated world capital stock (consisting of manufacturers) that Southern and Northern capitalists wish to put to work in the Southern primary production at any given point in time. That decision in turn is guided by the fundamental requirement that capital earns the same return wherever employed. Instantaneous North and South profit rate equalization together with real wage fixity in the South yields one of the model’s most crucial and distinctive analytical features: A one-to-one inverse relationship in the short term between the level of the terms of trade and the size of the capital stock engaged in the Northern manufacturing.

Overall, short-term equilibrium is established when world markets for capital and the two regions’ outputs clear at predetermined levels of capital stocks regionally owned and of the Northern labour supply. Owing to saving and investments by the capitalists in both regions, capital stock increases over time along with the exogenously growing Northern and the passively adjusting Southern labour supply. As a consequence, the world economy is propelled, provided it is dynamically stable, through a succession of short-term equilibriums to a long-term steady state in which capital and labour will be growing at the same rate in both regions, per capita incomes and capital stocks remaining constant.

Like Findlay, Burgastaller & Saaverdra-Rivano conclude that from a long-term growth perspective, full participation in the world economy through free trade and capital mobility may entail severe costs for a Southern labour surplus region. As the Northern capital moves to the primary sector in the South, where the capital-labour ratio is uniquely fixed by subsistence wage level, the per-worker output will not be altered in the long run. However, part of that output will now be produced by and paid out to foreign capital, resulting in a decline in the per-worker Southern income, while the opposite will be true for the North. Thus capital mobility, in consequence, leads to long-term worsening of the per-
worker income gap between the South and the North. Similarly, technology progress in primary goods production in the South results in a decline in the Southern per-worker income, as well as in a deterioration in the South’s total income position relative to the North. Moreover, part of the South’s capital stock is owned by the North, which causes repatriation of earnings at the expense of the South.

One of the most important pieces of empirical work in this area was by Chenery & Strout (1966). In their paper, they sought to develop a theoretical framework to analyze the process of development with external assistance in quantitative terms. This framework was then used to evaluate the performance of developing countries as well as evaluate their future needs for assistance under various assumptions. In their model, they considered underdeveloped countries characterized by limited structural flexibility. Using the Harrold-Domar growth model (Solow 1950), they considered the role of aid in the transition, focusing on both investment-limited growth and trade-limited growth. In their analysis, they concluded that the effectiveness of external resources in the short term was dependent on their use to relieve shortage of skills, savings, and imported commodities. However, even if the short-term productivity of aid is high, the economy may continue to be dependent on external assistance indefinitely unless the additional output is allocated so as to increase savings and reduce the trade gap. Over the whole period of transition to self-sustained growth, the use that is made of successive increments in output is likely to be more important than the efficiency with which external assistance is utilized in the first instance.

Similarly, Papanek (1973) conducted a cross-country regression partial analysis using 34 countries for the 1950s and 51 countries for the 1960s, examining the effects of foreign aid, FDI, and savings on economic growth. The study concluded that savings and foreign inflows explained about one third of growth, which was consistent with the history and analysis of development. On the determinants of FDI, Agarwal (1980) reviewed several studies that examined FDI as a function of the international differential rate of return on capital investment and observed that attempts to test this hypothesis statistically had failed to produce conclusive results. Few studies were found to either partially or wholly support the hypothesis, while many others could not find any association between the flow of FDI and international differences in returns on investment.

In sum, both Brems’s (1970) and Findlay’s (1978) models suggest that FDI in possession of superior technology can benefit by investing abroad to seek higher returns. In return, host countries can benefit from FDI, not only through increased capital accumulation but also through technology transfer. This is due to the existence of structural asymmetry between regions. However, Findlay observed that this is only possible in the short term. In the long term, foreign capital
rets the process of technology borrowing. Findlay (1980) and Burgastaller & Saaverdra-Rivano (1984) further show that structural asymmetries between the North and the South are pervasively biased against the latter. Hence, both the terms of trade and capital mobility hurt the South in the long run. These models, however, do not capture the entire complexity of FDI and other forms of international production. As a matter of fact, they explain only money capital, and within the models this is not clearly distinguished from portfolio flows. In addition, investors do not always seek to maximize profits, and there may be other short-term and long-term reasons why firms may invest abroad.

Turning to the assumption of perfect markets underlying these models, Lucas (1990) has observed that if the world capital markets were free and complete, it is clear that, in the face of return differentials of this magnitude, investments would flow rapidly from the US and other wealthier countries to India and other poor countries. Thus, the assumptions on technology and trade conditions inherent in these models are wrong. He observes that these models do not take into account cases of market failures, as indicated by their failure to account for differences in labour quality or human capital per worker. In addition, external benefits of workers’ and knowledge spillovers are assumed to be non-existent. Moreover, capital market imperfections present a need for an effective mechanism to enforce international agreements, especially arising from political risks. More importantly, in a world characterized by international free trade and perfect competition, FDI need not occur.

Findlay-type North-South growth and trade models have the principal feature of an exogenously given real wage in the South. As Chui et al. (2002) observe, this can be unrealistic for many types of North-South interactions. Indeed, it can be argued that real wage may be institutionally determined in some labour markets in the North, for example in Western Europe. This also applies to the Brems model, where money wages are assumed to be the same across sectors in each country and the prices of commodities within the sectors vary causing the real wage to vary. In addition, the notion of labour immobility in the North is unrealistic in many countries, especially with regional integration. Likewise, the labour surplus notion assumed in the North-South models may not be applicable to all countries in the South, especially East and South East Asia. Furthermore, some nations in the South are actively engaged in manufacturing, including exports of manufacturers, a fact which rests uncomfortably with the notion of surplus labour. It is also important to note that, with regard to capital mobility, there exist some capital flows from South to South and from South to North.
Business school models

The Business School or the Industrial Organization theories of FDI in general focus on the why, where, how, and when in understanding MNEs’ activities. The most dominant analytical framework, accommodating a variety of operationally testable economic theories of the determinants of FDI and foreign activities of MNEs, for the last three decades is the Dunning eclectic (or OLI) paradigm (Dunning, 2000). This framework avers that the extent, geography, and industrial composition of foreign production undertaken by MNEs is determined by the interaction of three sets of interdependent variables, which themselves comprise the components of three sub-paradigms. The first is the competitive advantages of the enterprises seeking to engage in FDI, advantages which are specific to the ownership of the investing enterprises. These he referred to as their ownership-specific (O) advantages. The second is the location attractions of alternative countries or regions for undertaking value-adding activities of the MNEs: The locational-specific (L) advantages. The third sub-paradigm of the OLI triad offers a framework for evaluating alternative ways in which firms may organize the creation and the exploitation of their core competencies, given the locational attractions of different countries and regions. These are the internalization (I) advantages.

The eclectic paradigm further asserts that the precise configuration of the OLI parameters facing a particular firm and the response of the firm to that configuration are strongly contextual. In particular, it will reflect the economic and political features of the country or region of the investing firms and of the country or region in which they are seeking to invest; the industry and the nature of the business activity in which the firms are engaged; the characteristics of the individual investing firms, including their objectives and related strategies; and the reason for the FDI. Regarding the last contextual variables, scholars have identified four main types of foreign-based MNE activity: Market-seeking, resource-seeking, rationalized or efficiency-seeking, and strategic-seeking FDI. The content and predictions of the eclectic paradigm are embedded in a number of different economic and business theories. Taken separately none of these offer a comprehensive explanation of the growth and decline of MNEs’ business activity, while taken together as a group they do. Most of the theories are complementary rather than replaceable by each other (Dunning 1998).

The theories of ownership advantages date back to the 1960s, pioneered by the work of Hymer (1960, 1976) and Kindleberger (1969), and are more concerned with why international production by MNEs takes place. They criticized the neoclassical approach for its limited ability to explain FDI flows. Their main argument was that the assumption in the neoclassical theory of perfect competition could not explain FDI, which in their view needed structural market imperfec-
tions to flourish. In a perfect market, all firms have equal access to the same resources and capabilities within countries, while there is complete immobility of resources and capabilities between countries, an assumption present in the Heckscher-Ohlin type of trade models. Potential sources of ownership advantages include product differentiation (imperfect goods markets), managerial expertise, new technologies or patents (imperfect factor markets), the existence of internal or external economies of scale, and government regulation. These advantages balance out the disadvantage of entering a foreign market (including higher risk, less information, more uncertainty, physical distance and differences in culture, business ethics, the legal system, and other regulations) in order to compete with local firms (Faeth 2009). The ownership sub-paradigm asserts that, *ceteris paribus*, the greater the competitive advantages of the investing firms relative to those of other firms, and particularly those domiciled in the host country, the more they are likely to engage in or increase their foreign production.

The locational advantage theories, on the other hand, are concerned with where and when international production takes place. For the most part, until recently few studies gave much attention to how the growth of the cross-border activities of firms might be explained by the kind of location-related theories which were initially designed to explain the siting of production within a nation state, or how the spatial dimension of FDI might affect the competitiveness of the investing entities. In the last decade, however, there has been a renaissance of interest by economists (Krugman 1991; Venables 1998) and industrial geographers (Storper 1995; Scott 1996) in the spatial concentration and clustering of some kinds of economic activity, by economists in the role of exchange rates in affecting the extent, geography, and timing of FDI (Frost & Stein 1991), and by business scholars (Enright 1991; Porter 1994) in the idea of an optimal locational portfolio of assets as a competitive advantage in its own right.

One of the earliest theories to focus on locational advantages was the product cycle theory (Vernon 1966), which held the view that investment decision was a choice between exporting and investing, as products moved through a life cycle divided into three stages (new, mature, and standardized products), giving a cost-based rationale for the switch from exporting to foreign-based production. Others include Knickerbocker (1973), who argued that MNEs were active in imperfect markets (oligopolistic) and invested as a result of a ‘follow-the-leader’ strategy or in reaction to foreign firms ‘invading’ their home market. This was one of the earliest attempts to explain geographical clustering of FDI. Similarly, Rugman (1979) developed the risk diversification theory, which suggested that MNEs normally prefer a geographical spread of their foreign investments to having ‘all their eggs in the same locational basket’. The location sub-paradigm asserts that the more immobile, natural, and created endowments – which firms need to use
jointly with their own competitive advantages – favour a presence in a foreign rather than a domestic location, the more the firms will choose to augment or exploit their O-specific advantages by engaging in FDI.

Lastly, the internalization theories are concerned with how international production takes place in the presence of O and L advantages. Orthodox internalization theories, first formalized by Buckley & Casson (1976), suggest that as long as the transaction and coordination costs of using external arm’s-length markets – in the exchange of intermediate products, information, technology, and marketing techniques, among others – exceed those incurred by internal hierarchies, then it will pay for the firm to engage in FDI rather than conclude a licensing or another market-related agreement with a foreign producer. In general, the transaction costs of using external markets tend to be positively correlated with the imperfections of those markets. The internalization sub-paradigm avows that the greater the net benefits of internalizing cross-border intermediate product markets, the more likely a firm will prefer to engage in foreign production itself, rather than license the right to do so (Dunning 2000).

Following the development of the industrial organization theories, numerous studies have been conducted analyzing the factors related to ownership, location, and internalization advantages through the eclectic paradigm or the sub-paradigms. Brouther et al. (1996) examined the impact of ownership and location variables on entry-mode selection in the US computer software industry. They found that as ownership and locational advantages increased, firms tended to utilize more integrated models of entry, supporting the Dunning framework. Similarly, Tse et al. (1997) analyzed the impact of OLI factors on model selection within China. They also found that as OLI advantages increased, firms tended to utilize more wholly owned modes of entry, again supporting the Dunning framework. Agarwal & Ramaswami (1992), in a more comprehensive study, examined not only the main effects of Dunning’s three type variables but also a set of six variable interactions: Firm size and experience with market potential, and investment risk; ownership advantages (ability to differentiate products) with investment risk, contractual risk, and market potential; and market potential with investment risk. Based on a study of 97 US equipment-leasing companies, they found that larger MNEs tended to use higher integrated (more wholly owned) entry modes in low-potential markets than did smaller, less-experienced firms. In a comprehensive literature review on the empirical studies arising from the eclectic paradigm, Faeth (2009) observes that the OLI framework does well in explaining FDI determinants.

Nevertheless these theories are not without their share of criticisms. First, the I theories are perceived to be incomplete in that they ignore other functions which a firm may perform aside from those related to transactions. In addition, by
stressing short-term profit maximization, they ignore other reasons why firms may wish to engage in business activities outside their national boundaries. Second, these theories are considered to be static in nature and give little guidance on how best a firm may organize its activities to create future assets rather than just optimize the use of existing assets. Third, the growth of a range of inter-firm coalitions has resulted in de facto internalization but without equity ownership, due to contractual agreements in cases of property rights and collaborative agreements for specific purposes such as a research and development project or a specific marketing strategy, among others (Dunning 2000).

The Dunning eclectic paradigm has also been criticized for several reasons. First, the explanatory variables identified in the paradigm are so numerous that the predictive value is almost zero. Second, there exists interdependence of the variables across the board. For example, a firm’s response to its exogenous location variables might itself influence ownership advantages, including its ability and willingness to internalize the markets. Third, the paradigm insufficiently allows for the differences in strategic response of firms for any given configuration of the OLI variables (Dunning 2001). Furthermore, it offers little guidance as to the dynamics of the internalization process of the firms or countries. Finally, Kojima (1982) criticized the OLI eclectic paradigm as well as other internalization scholars as being purely a microeconomic phenomenon.

Old growth and new trade models

The term ‘new’ theory in general refers to models with increasing returns to scale, product innovations, and product cycle (imitation by the South) as used by Chui et al. (2002). These models have emerged as a result of the integration of some of the aspects in the business school theories with trade theories and have contributed to the understanding of different forms of FDI. In addition, some of the new trade models have integrated with the growth theory to explain the North-South interactions. These growth models do not explicitly refer to the process of endogenous growth; hence the sobriquet, old growth and new trade theories. The innovative paper in this respect is by Krugman (1979). In this model, there are two types of goods, new and old. New commodities were innovated and produced exclusively in the North, while old goods can be manufactured in either region. The North is the innovating region as far as the new products are concerned, but there is a process of technology transfer (imitation) from North to South which is exogenous, akin to radioactive decay. Labour force is assumed to be the only factor of production. Dollar (1986) extended the Krugman model to incorporate two factors of production capital and labour, endogenized North-South technology transfer, and considered a process of international
capital mobility. Thus, this model preserves and builds on the features of the paper by Krugman and is here reviewed in brief.

In Dollar (1986), the world is divided into two regions, the North and the South, and there exist a large number of goods in the system, which can be produced using three factors of production: Capital, labour, and know-how. Capital and labour are assumed to be homogenous throughout the world, while the know-how to produce each good originates from the North. The defining characteristic of the region is the differential speed of factor mobility between the regions, compared with mobility within the regions. The know-how to produce a good diffuses instantaneously throughout the North but only becomes available in the South after a time lag. Thus, at one point in time all goods are either old goods, whose technology of production is available in both regions, or new goods that have been recently produced with new technology only available in the North.

The movement of capital between these two regions also takes place slowly over time, so that each movement of stock of capital in each region is fixed. Labour is assumed to be fixed even in the long term, while within each region there is instantaneous perfect mobility of capital and labour and zero transport costs. Once a region has the know-how to produce a good, it is assumed to be available within the region, such that only capital and labour are considered as inputs in the production function, which is assumed to be of the Solow type, with constant returns to scale. Moreover, the rate of transformation between the two types of goods is assumed to be different in the two regions, and the opportunity cost of a new good in terms of an old good is unitary in the North and infinite in the South. This creates a basis for trade. There is only one relative price of a new good in terms of the old one, and with the perfect competition assumption within each region the prices of all old goods must be the same, as it must for all the new goods.

The model further assumes equilibrium with complete specialization and that all the individuals within the two regions have the same utility function, which is borrowed from Dixit & Stiglitz (1977). Hence, two goods with the same price are consumed in the same quantities by all the consumers, and therefore all goods produced in the North, which must have the same price, are produced in the same quantity and at the same price in the South. This is considered representative of Northern and Southern goods and together comprises the total number of goods available in the world. Consumer welfare increases with the variety of goods consumed. On the supply side, given the assumption of perfect competition within regions, all goods produced in a region have the same price, which is equal to average cost. In addition, factor prices are expected to adjust so as to bring about full employment of resources in both regions. In a static model at equilibrium, each region exports some quantities of each of the goods it pro-
duces, and as a result there is a flow of new goods from the North to the South and a flow of old goods in the opposite direction.

Using the features in the Vernon (1966) product cycle model, the introduction of new products as a result of innovation increases the number of goods produced in the North. The transfer of technology, on the other hand, will change the new goods into old goods, reducing the number of goods produced in the North and increasing the number of goods produced in the South. The incentive for transferring production to the South in this case is that it can produce with the new technology more cheaply than the North. This difference in cost creates the potential for economic profits, which although never realized in the model, provides incentives for firms in the North to move production to the South (FDI), and for firms in the South to start imitating the northern products. The difference in the costs of production in the North and South is conveniently summed up in the terms of trade. As long as there is complete specialization, any good that the South learns to produce will cease to be produced in the North. Thus the process of innovation and technology transfer together determine how the ratio of the number of goods produced in each region changes over time.

In a dynamic environment, the North continually introduces new goods. This creates a steady flow of technology transfer to the South. For there to be an incentive for such a flow, the long-term equilibrium terms of trade must be greater than unitary so that the cost of production is lower in the South than in the North, for goods that can be produced in both regions. The fact that the terms of trade must be greater than unitary, while the returns to capital are equalized, means that wages must be higher in the North than in the South. That is, for there to be a difference in the cost of production, the internationally immobile factor of production must earn more in the North than in the South. Thus, the ability of the North to introduce new products and monopolize them temporarily enables the workers in the North to earn a premium over the wages of their counterparts in the South, even though labour is assumed to be of the same quality in both regions. Furthermore, the wage-labour ratio in the North will be higher than in the South, suggesting that less labour per unit is used for production in the North. Thus, although all goods are produced using the same production function, the North will be more capital-intensive, while the South will be more labour-intensive, as would be expected in the factor proportion model. However, the differences in relative factor endowment develop as an endogenous result in the model, owing to the North’s ability to introduce new products and not as a result of initial differences in factor endowments.

---

4 The model assumes that the rate at which new goods become old in the North is a positive function of the differences in the cost of production in the two regions such that the greater the potential economic profits to be gained through technology transfer, the more such transfers take place.
Finally, a rise in the effective and fully employed labour force in the South will eventually lead to a decline in the ratio of goods produced in the North relative to the South. Upon impact, the North’s terms of trade will improve. This will cause capital outflow to the South and a greater degree of technology transfer to that region. This implies an increase in the South’s competitiveness. This new equilibrium will see a decline in the North’s wages. It should be noted, however, that the crucial assumption in the model which causes the decline in Northern wages under full employment is the flow of capital from North to South, and not the manner in which new goods are innovated or knowledge spills over. An exogenous increase in the level of diffusion to the South leads to a rise in the South’s terms of trade, capital-labour ratio, and the absolute level of real wages. Nevertheless, the North continues enjoying higher wages relative to the South as the only innovating region. Changes in effective labour and the rate of imitation only narrow the North-South gap but do not eliminate it.

The Krugman (1979) and Dollar (1986) models, unlike Vernon’s (1966) product cycle model which assigns a central role to FDI, cast imitation as the only channel of international technology transfer from an innovating North to an imitating South. In addition, the rate at which an individual firm discovers and successfully markets new products is treated as exogenously given. This is tantamount to implying that from an individual firm’s perspective, successful product innovation is either effortless or guaranteed by large expenditure on new product development. This is obviously not the case. As Schumpeter (1942) observed, firms compete with each other to successfully produce new products. In undertaking research and development, there are losers and winners because firms can spend substantial resources on new products, only to find that another firm has discovered and patented the new product. In addition, the assumptions of zero transport costs and the South being only an imitator are not necessarily true. Transport cost matters, and in recent times the South is also involved in innovations. Diwan & Rodrik (1991) observed that the North and the South may have different technological needs, causing the South to develop patents based on cheap labour, to promote development of technologies appropriate to the South.

With regard to determinants of FDI, the neoclassical trade theories have integrated with the Business School theories to explain why MNEs locate to foreign countries. As a result, two important explanations have emerged. The first is based on the factor proportion hypothesis, while the second is based on the proximity-concentration hypothesis. The factor proportion hypothesis focuses on the vertical expansion characteristic of North-South flows (Helpman 1984; Markusen 1984; Helpman & Krugman 1985; Either & Horn 1990). A factor proportion model with differentiated products predicts that when factor proportions are identical, the differential in GDP shares of the source and destination countries
and their joint incomes should be the only determinants of trade volumes, and there should be no multinational activities. In addition, the model predicts that the intra-industry share of trade flows should be decreasing in factor proportions differences, while MNE activities should rise only in a single direction between countries with large factor proportions differences.

The proximity-concentration model, on the other hand, formulates location advantages in terms of proximity to customers or specialized suppliers, which motivate horizontal expansion across borders at the expense of reduced scale (Krugman 1983; Horstmann & Markusen 1992; Helpman et al. 2004). This model predicts that MNE activities will arise between countries with high transport costs and trade barriers in industries with low plant scale economies and involve increasing two-way activity the more similar factor proportions are. The factor proportion model further predicts that all varieties of a final good produced by a foreign affiliate should be exported back to the headquarters market, while in the proximity-concentration hypothesis, MNEs will substitute overseas production for trade in final goods, so that flows of final goods back to the parent company will be zero. Within a multi-stage production process, the factor proportions model predicts that MNEs will generate one-way trade in intermediates, while the proximity-concentration model predicts they will generate two-way trade in intermediates.

Brainard (1993) tested the validity of the factor proportions hypothesis using the US data. The study found some support for the hypothesis, when the affiliate production destined for export to the parent’s market was compared with that destined for sale in the local market. The former was found to be more responsive to factor proportion differences. However, the two types of activity differed more in their responses to transport costs and destination market income. The study found that only a small part of MNE activity in and out of the US in the late 1980s could be explained by factor proportion differences. Brainard (1997), using the same data, found that as overseas production by MNEs increases relative to exports, the higher become the transport costs and trade barriers and the lower become the investment barriers and scale economies at the plant level, relative to the corporate level, thus supporting the proximity-concentration hypothesis. Ekholm (1998), using data on Swedish and US MNEs, found similar results. Helpman et al. (2004), while also confirming this hypothesis and using US export and affiliate sales data covering 52 manufacturing sectors and 38 countries, showed that cross-sectoral differences in firm heterogeneity predict the composition of trade and investment. These empirical studies confirm Faeth’s (2009) observations in a comprehensive literature review on the factor proportions and proximity-concentration hypothesis, where evidence on the latter was
found to be robust. Market size, transport costs, and trade barriers increased FDI, while factor endowments were only relevant in some cases.

New growth and new trade models

The term ‘new growth theory’ is associated with a body of empirical work that was developed in the 1980s, which distinguishes itself from the neoclassical growth theory by emphasizing that economic growth is an endogenous outcome of an economic system, not the result of forces that impinge from outside. For this reason, this theoretical work does not invoke exogenous technological change to explain why income per capita has increased by an order of magnitude since the Industrial Revolution. In addition, it does not assume that technology is the same across countries. The new growth theory has developed along two principal distinct paths: Increasing varieties and quality ladders. The former was initiated by Romer (1986) and focuses on the dynamic externalities within a perfect competition framework. Similarly, Lucas (1988) developed a growth model in which human capital is taken to be specific to the production of particular goods and is acquired on the job or through learning by doing. However, both works are not dissimilar from the theme of learning by doing, which has been present in development literature for some time (Bardhan 1995). The main contribution to new growth theory is based on Romer (1990). Within this paradigm, imperfect competition – and with it monopoly profits – is what motivates the innovator. The resulting allocation of labour between manufacturing and innovation has provided new insights both for the closed and the open economy. Peretto (1999) has extended this approach to show how developing economies can go through an industrial cycle involving accumulation of physical and human capital, followed later by research and development effort.

The quality ladders growth models are based on the argument that technology progress occurs in the form of innovations, which in turn result from the activities of research firms. As a departure from the Romer-type models, they emphasize that technology progress creates winners and losers as well as gains, by rendering obsolete old skills, goods, markets, and manufacturing processes. Each innovation consists of a new line of intermediate goods that can be used to produce final output more efficiently than before. Research firms are motivated by the prospects of monopoly rents that can be captured when a successful innovation is patented. But those rents in turn will be destroyed by the next innovation, which renders obsolete the existing line of intermediate goods (Segerstrom et al. 1990; Grossman & Helpman 1991a; Aghion & Howitt 1992).

Glass & Saggi (1998), building on a quality ladders product cycle model, explored how the quality of technology transferred through FDI is linked to innova-
tion and imitation when the absorptive capacity of LDCs is limited. Successful imitation of low quality levels makes FDI involving high quality levels possible, through the reduction of the technology gap. A subsidy to imitation or a tax on low quality FDI production encourages imitation relative to innovation, thus releasing the constraint faced by foreign firms to produce in the South. These forces stimulate high-quality FDI and raise Southern welfare through lower prices and increased innovations and wages. The quality of ladder growth models are easily extended to a trade framework with two equivalent countries under certain assumptions, and the results are identical to those of increasing varieties models.

One of the most popular North-South product cycle models for an open economy was developed by Grossman & Helpman (1991), where new innovative goods are developed and produced in the North and exported to other countries, North or South. Eventually, the South develops the ability to imitate many of these activities, and production shifts to this region. But trade between nations engaged in producing innovative goods in the North and the imitating South clearly do not exhaust all the possible trading patterns. Chui et al. (2001) observe another intermediate phase of development: Innovation in the North and innovation coexisting alongside imitation in the South. In their model, they allow for a copying phase as one of the several possible stages of development for the South. In addition, they introduce a second factor of production – skilled labour alongside a high technology sector, producing an expanding variety of differentiated goods – and include a ‘traditional sector’ with low technology levels in both regions. Considering the output of the high technology sector as one good, these changes give a model a two-factor, two-goods, Heckscher-Ohlin structure, suitable for accessing free trade. The South is allowed to differ from the North in three ways. First, it is less efficient at adopting the new technology available on a worldwide level. Second, the speed with which the South learns from the North is less than in the opposite direction. Third, the endowment of skilled labour is less in the South than in the North.

These three asymmetries, together with the degree of patent protection enjoyed by the Northern firms, give rise to four stages of development in the South as possible equilibria. *Ceteris paribus*, an increase in the North-South knowledge diffusion rate enables the South to progress to a higher state. An increase in either Southern inefficiency or international patent protection has two effects, depending on the speed of the North-South knowledge diffusion. For a low level of knowledge diffusion, it enforces the South’s specialization in the traditional

---

5 Less Developed Countries (LDCs) are countries whose state of economic development is characterised by low national income, a high rate of population growth and unemployment, and dependence on commodity export.
good. A high level of knowledge diffusion encourages innovation in the South at the expense of copying. An increase in the proportion of skilled workers in the South favours a progression into a higher stage of development, irrespective of the rate of knowledge diffusion. In this model, trading between countries results in workers in the North gaining, while only unskilled workers in the South gain. However, if the international diffusion of knowledge capital from North to South is sufficiently rapid, then even skilled workers in the South can benefit from trade. Thus, according to this model, policies that can promote world integration may increase world growth, first through specialization in which the North devotes more resources to innovative research and development, and second through knowledge spillovers which enable the South to move to higher stages of development.

Several studies have been conducted empirically testing the validity of these models, with extensions being made to incorporate FDI (by assuming that knowledge is a function of FDI), and contributing to the understanding of its impact on economic growth. Borensztein et al. (1998) tested for the effect of FDI on economic growth in a cross-country regression framework, utilizing data from industrial countries to 69 developing countries since the 1970s. MNEs were assumed to possess more advanced ‘knowledge’, which allowed them to introduce new capital goods at lower costs. However, the application of these more advanced technologies required the presence of a sufficient level of human capital in the host economy. The stock of human capital in the host economy therefore limited the absorptive capacity of a developing country. They found that FDI is an important vehicle for the transfer of technology, contributing to growth in larger measure than domestic investment. Moreover, they found a strong complementarity between FDI and human capital – that is, the contribution of FDI to economic growth is enhanced by its interaction with the level of human capital in the host economy. However, their empirical results imply that FDI is more productive than domestic investment only when the host country has a minimum threshold of human capital. In addition, they observed that FDI crowds in domestic investment. They therefore concluded that the main channel through which FDI contributes to economic growth is by stimulating technological progress, rather than by increasing total capital accumulation in the host economy.

Similarly, de Mello (1999) estimated the impact of FDI on capital accumulation, output, and total factor productivity growth in the recipient economies using a sample of OECD and non-OECD countries in the period 1970-1990. Applying both time series and panel data evidence, he found that the extent to which FDI is growth-enhancing is dependent on the degree of complementarity and substitution between FDI and domestic investment. While the former promotes growth, the latter does not. In a survey of literature on the impact of FDI on growth of
developing countries, de Mello (1997) also found that the ultimate impact of FDI on output in the recipient economy depends on the scope of efficiency spillovers to domestic firms, by which FDI leads to increasing returns in domestic production and increases in the value-added content of FDI-related production. Car-kovic & Levine (2005), using two datasets from the World Bank and IMF and the Generalized Method of Moments (GMM) panel estimator, reassessed the relationship between economic growth and FDI. They found that the exogenous component of FDI does not exert a positive robust influence on economic growth.

There have been numerous microeconomic studies examining the spillover effects of MNEs’ activities, and over time different scholars have attempted to take stock of the evidence. Beginning with Blomstrom & Kokko (1998), scholars have observed that FDI can result in benefits for host countries even if the MNEs decide to carry out their foreign operations in wholly owned affiliates, since technology to some extent is a public good. These benefits arise in the form of externalities, which they identified as productivity spillovers and market access spillovers. They found evidence of spillovers from MNEs, which in their view may be substantial both within and without industries. However, they find no strong evidence on their exact nature and magnitude. In addition, they observed that host-country spillovers vary systematically between countries and industries, and the positive effects of FDI are likely to increase with the level of local capacity and competition. Likewise, Gorg & Strobl (2001) found that cross-sectional studies find more significant evidence of positive knowledge spillovers than panel studies do, suggesting the presence of unobserved firm heterogeneity often not controlled for in the former studies.

Gorg & Greenaway (2004) surveyed more than 40 econometric studies mainly micro in nature, and found that empirical evidence is at best mixed, although the majority of the studies suggested the presence of positive spillovers or insignificant results. Saggi (2002) in a literature review on technology transfer also found that macro studies provide positive evidence of the impact of FDI on growth, while plant-level studies do not find evidence of productivity spillovers. Smeets (2008) too, in a literature review on knowledge spillovers found that studies that

---

6 Productivity spillovers are said to take place when the entry or presence of an MNE affiliate leads to productivity or efficiency benefits in the host country’s local firms, and the MNEs are not able to internalize the full value of those benefits. Another kind of productivity spillover occurs if the entry of an affiliate leads to more severe competition in the host economy, such that local firms are forced to use the existing technology and resources more efficiently. A third type of spillover effect takes place if competition forces local firms to search for new, more efficient technologies.

7 MNEs often possess strong competitive advantages in entering world markets, such as experience and knowledge of international marketing, established international distribution networks, and lobbying power in their home countries. As a result of their own export operations, MNEs may pave the way for local firms to enter the same export markets, either because they create transport infrastructure or because they disseminate information about foreign markets that can be used also by local firms.
take into account individual spillover channels find robust evidence of knowledge spillovers from FDI. In addition, studies on the importance of mediating factors and FDI heterogeneity were found to be less conclusive, while in general most studies did not distinguish properly between knowledge spillovers and transfer.

More recently, Mebratie & Bergeijk(2012) took stock of the literature on FDI spillovers. They analysed econometric studies on FDI spillover effects that were published in the period 1983-2008. Thirty of these studies were national, representing developing and emerging countries. They found that these studies tend to ignore two sources of heterogeneity: Exports and R&D. Using a meta-analysis to correct for the differences in research design (including regional effects, sample size and level of aggregation), they investigated the spillover effects of foreign firms and domestic firms. They show that developing and emerging countries exhibit a lot of variation in the extent of heterogeneity of their populations of firms. In addition, they show various effects of different measures of heterogeneity. Firm size, labour quality and exports are found to have more positive effects for the development of firms compared to that of foreign ownership. Domestic firms could also be affected negatively by foreign firms due to the adverse effects of competition in product and resource markets. They thus conclude that the developing opportunity of local firms in developing countries is faster if they focus on improving production capacity and produce for the international market.

On FDI determinants, there have been new developments within the new trade theories, aimed at integrating the factor proportions and proximity-concentration theories, allowing a firm the options of building multi-plants or geographically separating headquarters from a single plant (Markusen 1997, 2002). This approach has generated the knowledge-capital model based on the assumption that knowledge is geographically mobile and also a joint input to multiple production facilities. Within this model, various combinations of vertical (inter-industry) MNEs and horizontal (intra-industry) MNEs and strictly national firms can arise endogenously as a function of parameter values given as trade costs, differences between countries in relative and in absolute factor endowments, and investment barriers. As a consequence, Markusen & Maskus (2002) observe that there is now a reasonably well-developed set of theories that have implications for the relationship between MNE activity and country characteristics such as size and relative endowments. Arising from these models is the conclusion that vertical MNEs dominate when countries are very different in relative factor endowments, while horizontal MNEs dominate when countries are similar in size and in relative endowments and when trade costs are moderately high. In some cases, foreign investment and trade liberalization lead to a reversal in the direction of trade. Investment liberalization can also lead to an increase in the volume of trade
and produce a strong tendency towards factor price equalization. Hence, these models conclude that FDI can be a complement to trade both in volume and in a welfare sense.

The empirical work in this area is relatively new. Carr et al. (2001) were the first authors to estimate the knowledge-capital model, using a panel of inward and outward sales data of foreign affiliates for the US and 36 other countries. They tested for the importance of MNE activity between countries as a function of certain country characteristics, in particular size, size differences, relative endowment differences, trade and investment costs, and certain interactions among these variables as predicted by theory. They found that outward investment from a source country to affiliates in a host country was increasing in accordance with the sum of the countries’ economic sizes, their similarity in size, the relative skilled labour abundance of the parent nation, and the interaction between the size and relative endowments. These results were consistent with earlier findings by Brainard (1993, 1997) and Ekholm (1998). These findings were, however, disputed by Blonigen et al. (2003), who claimed that the Carr et al. empirical work (2001) mis-specified the variables measuring differences in skilled labour abundance. Taking this into account, they estimated a model whose results supported the horizontal model and not the knowledge-capital model, as MNE activity was smaller the more countries differed in their relative endowments.

Markusen & Maskus (2002) also analyzed the importance of the horizontal, vertical, and knowledge-capital FDI models in explaining the overall pattern of world FDI activity. Both the horizontal and the knowledge-capital models were found to be more descriptive but almost indistinguishable in data, while the vertical FDI model had little explanatory power and could not explain aggregate world FDI. This result, however, did not reject the theory that vertical models were important for some industries in some host countries. Davies (2008) observed that one of the reasons for different conclusions in this literature may be due to the extent to which unaccounted fixed effects’ biases vary across different datasets. Taking this into account, Mariel et al. (2009) found that the vertical component of the knowledge-capital model was relevant. On the issue of the relationship between FDI and trade, Camarero & Tamarit (2004) estimated the demand for exports and imports of manufactured goods for a panel containing the majority of EU countries as well as the US and Japan. Their results pointed to the complementarily between trade and FDI.

Recently, new studies have acknowledged the existence of hybrid or ‘complex’ MNEs, which are neither purely horizontal nor purely vertical (Yeaple 2003; Eggar et al. 2004; Grossman et al. 2006; Ekholm et al. 2007). This, as well as the interest in more complex integration strategies, has initiated departures from the two-country case, putting emphasis on the role of endowments and trade
and investment costs all over the world. This has resulted in the inclusion of third-country effects in the knowledge-capital models, since the average country pair is relatively small compared with the rest of the world. The strength of third-country effects depend on a host country’s relative remoteness from third markets and can affect bilateral FDI owing to their weight in worldwide demand and supply. In addition, from an econometric point of view, failure to include third-country effects leads to biased estimates of the determinants of bilateral FDI. Baltagi et al. (2007), using a panel dataset on US industries and host countries observed over the 1989-1999 period, estimated a ‘complex FDI’ version of the knowledge-capital model. They found that bilateral and third-country effects of changes in skilled and unskilled labour endowments tend to be substitutes for vertical and complex vertical FDI.

The uneven development positions

Theories of unequal exchange (imperialism) in the early twentieth century were presented by Hobson (1902) and Lenin (1916). Both theories, influenced by Marx’s writings, were essentially concerned with explaining the spectacular growth of the Western colonial empires in the late nineteenth century. While acknowledging the presence of political factors among others factors in this sudden colonial expansion, both theories specifically stressed the need for the Western economies to export capital and the subsequent need to protect capital overseas by force, as the primary engine behind the growth of the empires of the great powers. According to Lenin, the more developed the country, the lower the rate of profit and the greater the ‘overproduction’ of capital, and, consequently, the lower the demand for capital and the stronger the expulsion process. Hobson, on the other hand, also thought that excessive savings and inadequate consumption in the core countries was the taproot of imperialism. These problems were arising owing to the existence of labour constraints in the developed countries. Confronted by frequent recessions in the late nineteenth century, capitalists became ‘tempted’ more and more to use some distant undeveloped country by annexation and protection (Oneal & Oneal 1988; Eckstein 1991).

In the periphery (South), capital was scarce, labour abundant and unorganized, and markets could be shaped by economic and political power. In addition, many expenses for social overheads, administration, and the maintenance of order were borne by the colonized people or the imperial nation as a whole, not just the bourgeoisie. For this reason, it was believed that ‘super profits’ in the periphery were the principle attraction of empire, though access to natural resources, the control of trade, and feudal motives were also considered important (Oneal & Oneal, 1988). With the dissolution of the formal empires, MNEs are now perceived by dependency theorists as the primary tool in explaining neo-colo-
nialism. These MNEs dominate the primary and extractive industries in the South, resulting in a form of incorporation into the international system that tends to inhibit industrialization and relegate these countries to less dynamic forms of growth (dos Santos 1970).

Others like Cardoso (1972), however, saw the clear evidence that the manufacturing sector was expanding in the South, but noted that it was dominated by MNEs with their headquarters in North, which he considered neo-colonial. These corporations, despite whatever benefits they may bring in the form of managerial and technical know-how, take more than they give, and – what is more important – they make it virtually impossible for local, self-sustaining industrialization to occur. Indeed, most contemporary radicals (dependency theorists) believed that transnational capitalism is so exploitative in the periphery that contrary to Marx, Lenin, & Hobson, capital on balance is transferred to the core countries and development in the periphery is retarded (Baran & Sweezy 1966). In explaining capitalists’ excessive profits, Frank (1981) has stressed the ‘super exploitation’ of labour in the developed countries. Similarly, for Amin (1977) the key is that the differences in wages between the core and the periphery are greater than the difference in productivity, which ‘makes possible a higher overall rate of profit in the periphery’.

Apart from the dependency that the international system creates through MNEs in the South, it also leads to political elites in these economies who create structures for domestic rule on a coalition of internal interests favourable to international connections. Thus, it is not the sheer economic might of the outside that dictates the dependent status of the South, but the sociological consequences of this power. As a consequence, a symbiotic relationship has emerged over time, in which the system has created its servants, whose needs dictate that its survival be ensured whatever the short-term conflicts of interests may be (Frank 1981). In addition to these effects, MNEs operating in the South may have inadequate safety standards, employ child labour, create sweatshops in their factories, and pollute the host-country environment. Moreover, developing countries wishing to attract and retain MNEs and FDI are thus forced to participate in a global ‘race to the bottom’, where the country with the lowest standards receives a greater proportion of investment (Oetzel & Doh 2009). Thus, contemporary imperialism is thought to provide MNEs with outlets for capital that yield higher returns than in the North, at the expense of development in the South, and the only remedy in this situation is thought to be revolution and the creation of a new, self-reliant, socialist state.

In an alternative perspective, some scholars within the literature of unequal exchange have argued that investors are apt to avoid the periphery because returns there are not commensurate with the risks. These scholars, commonly re-
ferred to as ‘structuralists’, suggest that market forces tend to accentuate international inequalities. In contrast to the dependency theorists, who stress the value of cheap labour, the structuralists emphasize the attraction for investors of large, prosperous markets and favourable political and cultural environments. Thus, while the dependency theories emphasize the advantages of investing in the South, structuralists have noted the drawbacks: Underdeveloped markets, limited availability of appropriate skilled labour, inadequate infrastructure, limited economies of scale, uncertain legal systems, and unstable governments (Blaug 1972). As a result most governments in the North have subsidized investment in the South. Nevertheless, structuralists fear that the South is more subject to neglect than exploitation.

Most of the ideas within the structuralists’ view are based on the work of Prebisch (1950) and Singer (1950). They were both concerned about the rising gap between the North and the South and its relationship to trade, and they argued that international specialization along the lines of comparative advantage had excluded the South from the fruits of technical progress, which benefited the North. This was because the South was highly specialized in the production of export primary commodities, while technology was concentrated mainly in industry, and the relative prices of primary commodities in terms of manufactures had fallen steadily since the start of the nineteenth century. According to Prebisch, the downward trend occurred and was likely to continue because of strong labour unions in the North, which caused wages in manufacturing to ratchet upwards with each business cycle. This, in turn, ratchets up the cost of manufacturers. Hence, owing to weak trade unions in the South, the cost of primary commodities rises by less than that of manufactured commodities during upswings and falls by more during downswings, creating a continuous decline in the relative cost of primary commodities.

Singer (1950), on the other hand, argued that this trend was due to the monopoly power of manufacturers mainly controlled by MNEs, which prevented the technical progress in that sector from lowering prices. However, he also observed that the demand for primary commodities showed relatively low income elasticity, so income growth tended to lower the relative demand for and hence the price of primary commodities. Moreover, he argued that technical progress in manufacturing tended to be raw-material saving, thereby causing the demand for primary products to grow more slowly than for manufactured commodities, while also reducing employment in such sectors. Payoffs to elites in the South consist of dominant control of political institutions, economic entities such as franchises,

---

8 This was based on the observation that wages rise during the upswings, but are sticky during downswings.
and major landholdings. MNEs also attract local capital that further retards capital formation for some less advanced sectors of the domestic economy.

Based on these observations, the structuralists conclude as policy implications for developing countries that the only way out of this dilemma is for the developing countries to foster industrialization aimed at changing the pattern of comparative advantage. As a result, the adoption of import substitution policies by developing countries was encouraged. Through import substitution the South protects its domestic industries against imports, subsidizes domestic industries, particularly in manufacturing – including seeking FDI for such purposes – and enters into preferential trade agreements with other Southern countries. Import substitution was seen as most desirable to small countries with mainly agrarian or natural resource sectors that faced declining terms of trade with the larger countries in the North. The import substitution policy was adopted by many developing countries in the 1950s, but the disappointing performance of the state-led industrialization and the collapse of centrally planned economies led to the repudiation of these early models of development in the 1980s and 1990s.

Arising from these recommendations, several scholars have explored various strategies for industrialization. Rosenstein-Rodan (1943) held the view that the South required planned industrialization comprising several complementary industries. However, individual firms could not undertake such a massive investment since the demand for their commodities would not be guaranteed, owing not only to consumers’ love of variety, but also the free rider problem. Hence, while external funding was needed, MNEs provided additional support, owing to the technology transfer to the South. In addition, MNEs were preferred because capital borrowing was considered more expensive due to the risk in the South, both political and otherwise. These structural problems made the state constitute the sine qua non condition of international investment on a large enough scale. Thus active participation of the state in economic life was considered important if this big push towards industrialization was to be achieved. Murphy et al. (1989) formalized the idea of coordinated investment in the ‘big push’ theory, as given by Rosenstein-Rodan (1943) to understand the importance of demand spillovers between sectors, by looking at simple stylized models of less developed economies in which these spillovers were strong enough to generate the big push. They found that the big push was possible in economies in which industrialized firms capture in their profits only a fraction of the total contribution of their investment to the profits of other industrializing firms. In this case, firms can help foster a mutually profitable big push, even while an individual firm would lose money industrializing alone. This once again suggests renewed emphasis on the role of government in the development process.
Similar views were largely held by other structuralists (Lewis 1954; Myrdal 1957; Hirchman 1958) although with slight variations. Nurkes (1953) further argued that underdevelopment persisted owing to the existence of the vicious circle of poverty or the poverty trap. This was caused on the one hand by a thin domestic market due to low incomes, and on the other hand the supply of goods was scarce due to the inability of people to save. Thus, the level of capital accumulation, investment, and productivity was low. Under such circumstances, he felt that the free market was unable to direct capital towards socially efficient investment projects. Nurkes favoured a uniform industrial policy (i.e. the doctrine of balanced growth), which required a massive investment effort. In contrast, Hirshman (1958) maintained that countries in the South lacked managerial and entrepreneurial abilities. Therefore, the optimal policy should have a goal of unbalanced development, concentrating investment in those sectors with significant external effects, which can facilitate the promotion of complementary investments in the rest of the economy. Similarly, Rostow (1960) was also of the view that countries could emerge out of stagnation into self-sustaining growth as a result of aid finance which increases investment, leading to an economic take-off.

Most elaborations of the structural aspects of the economy and surplus labour concepts have been mainly made in a two-sector neoclassical system. However, empirical application of these models has been limited by lack of data on different aspects of dualism. Krugman (1981) developed a two-country model of capital accumulation and growth, where the industrial sector exhibits increasing returns to scale. In this model, it was shown that uneven development was a result of initial discrepancy in capital-labour ratios between two countries, which accumulates over time, leading to the division of the world into capital-rich industrial regions and capital-poor agricultural regions, resulting in lack of industrialization in the South. With international capital mobility, the model gave rise to a two-stage pattern of development. In the first stage, trade was the engine of growth in the lending country, while in the second, foreign investment takes the role which was reminiscent of the Hobson-Lenin hypotheses. Similarly, Findlay (1980) and Burgastaller & Saaverdra-Rivano (1984), discussed earlier, further suggest that trade and capital mobility from the North leaves the South worse off as suggested by the Prebisch-Singer hypothesis. However, some North-South models have challenged these observations. For example, Dollar (1986) and Chui et al. (2001) observe that international trade and technology transfer can benefit the South as long as it continues to enjoy a cost advantage, reflected in its terms of trade.

More recently, several scholars have continued to lay stress on the importance of government involvement in the industrialization process. However, as a depa-
ture from the traditional rationale for selective industrial policy in terms of market failures, there is a new emphasis on a broader approach to industrial development (Lall 2004; Rodrik 2004). They perceive industrial policy as a discovery process where firms and governments learn about underlying costs and opportunities and engage in strategic coordination. By so doing, they present a justification for industrial policy in the case of information externalities and technology externalities, in addition to coordination externalities arising from the big-push theory. Information spillovers are considered to arise when entrepreneurs cannot diversify their product structure since they lack information on new activities that can be produced at costs low enough to be profitable. This calls for subsidy in new, non-traditional industries. Technology externalities have to do with how enterprises in developing countries actually use technology. This school further argues that industrial success in developing countries depends essentially on how an enterprise manages the process of mastering, adapting, and improving upon existing technology. This process is considered difficult and prone to widespread market and institutional failures, with important implications for policy (Lall 1992, 2004).

Conclusion

Until the mid-1980s, growth theory assumed that economic growth and technical change were determined exogenously. According to this theory, policy affects the rate at which economies converge to the long-term (steady state) growth rate, but not the long-term growth rate itself. These models are referred to as old growth in this paper. Growth theory underwent a fundamental change in the mid-1980s with the development of the endogenous growth theory. The implication of the new growth theory is that the returns to accumulation of knowledge and human capital do not diminish at the aggregate level, because of positive spillovers effects, and that policies can have a permanent impact on economic growth. It can also be observed that within growth theories, some have concentrated on closed economies while others have explored growth in open-economy settings. The basic idea in these models is that goods embody technological know-how, and therefore countries can acquire foreign knowledge through imports. Both old and new growth models have integrated with trade theories and Business School theories in order to explain growth, international trade, and MNEs’ activities in North-South interactions. An obvious weakness of most of these models is that they deal with capital mobility in general, while only a few expressly address FDI.

The existence of distinct types of North-South models naturally reflects the different forms of North-South interactions, pattern of trade, in some cases different forms of FDI, and stages of development. The notion of surplus labour in
the Findlay-type models is more applicable in Southern countries at an earlier stage of development, especially within Sub-Saharan Africa, than is now witnessed in East and South East Asia. The models of Krugman & Dollar do not rest on asymmetries in the labour market but instead emphasize the absorption by the South of exogenous technical progress in the North. In the models of Grossman & Helpman developed by Chui et al, technical progress is endogenised, after the new growth models. Based on these models, two mechanisms have been identified by which openness may increase world growth and welfare of both the North and South, but not necessarily. The first is through specialization in which trade sees the North devoting more resources to innovative research and development. The second is through the knowledge spillovers which enable the South to progress into higher stages of development. However, this second channel has an ambiguous effect on world growth. In the stage where the South only copies, increased spillovers actually reduce the incentive to innovate in the North and the long-term growth falls; but as spillovers increase further, the South enters stages where it begins to innovate and world growth increases. This provides a case for international property rights, which have the effect of encouraging innovation and discouraging copying in the South. However, trade openness yields higher growth rates and welfare than autarky regimes even when the South is in its copying stage of development.

Since most of these models do not distinguish between the various forms of capital flows, it is difficult to account for the effects of FDI within the North-South interactions. By so doing, they assume that technology and research and development activities are costless, which is incorrect. Yet we know that MNEs account for much of the trade, innovation, and technology transfer from the North to the South. In addition, growth models based on the product cycle lay stress on imitation as an important channel of technology transfer; Vernon (1966) did stress the role of FDI as an important channel of international technology transfer. This reality was considered much by the early growth theories, which attempted to incorporate FDI within their growth models. In Brems’s case, although considering a North-North case, foreign firms with superior technology not only contributed to accumulation of capital, but also transferred technology in the host country. A similar view is held by Findlay, who further saw FDI generating technology transfer through contagion as foreign firms interacted with domestic firms within the host countries. In the endogenous growth models, FDI is incorporated in the models by assuming that knowledge production is dependent on FDI. As a result, FDI not only contributes to capital accumulation and technology transfer in the host countries but also plays an additional role of generating both productivity and market-access spillovers.
This neoclassical view of the potential benefits of FDI to the South has been challenged both by dependency theorists and the structuralists. The latter suggest that, owing to the structural rigidities existing within the South, the benefits of FDI cannot be realized. This is especially true in some of the countries in the South which only manage to attract FDI in the traditional sectors. Most of these countries continue to export primary commodities to the North while importing industrial goods from them, a phenomenon present in the Findlay models. As a result, they continue to experience unfavourable terms of trade as expressed through the Prebisch-Singer hypothesis. In addition, most of these economies have a small industrial sector, which remains uncompetitive. One of the policy dilemmas faced by these countries is whether they should go back to import substitution strategies to encourage industrialization once more or liberalize their economies, as suggested by the neoliberals, to encourage trade and FDI, which as noted earlier may contribute to some challenges, including unfavourable terms of trade. In the last decade, based on the experience of the newly industrialized countries, the structuralists have argued a case for policy intervention in the South, especially at the early stages of development, in order to develop absorption capabilities and institutions necessary to realize the benefits from trade and FDI. Some of these issues have been incorporated within the new North-South models of growth and trade. The dependency theories, which possess a more radical view about the plight facing the South, have become increasingly unpopular over the years, especially with globalization and the recent industrialization of countries in Asia.

On determinants of FDI, the literature has made great strides in unearthing the FDI motives. The classical and neoclassical theories assuming perfect competition have been unable to explain the occurrence of FDI. As many scholars have noted, the Hecksher-Ohlin model’s assumption of imperfect immobility of factors of production makes it impossible for FDI to arise. Even after relaxing this assumption, the traditional trade theories fail to account for complexities in FDI. As Hymer and others have noted, FDI is usually conducted by MNEs which possess unique advantages that allow them to compete with domestic firms within the host countries, and therefore FDI arises owing to the existence of market failures. In considering the determinants of FDI, one could consider advantages that are either internal to the firm or external to the firm. The advantages that firms possess which can enable them to compete in host countries are those of ownership and internalization, discussed in the Dunning eclectic paradigm, which includes new technology, products, methods of production, economies of scale, knowledge of markets, managerial capabilities, and organization abilities.

The advantages that are external to the firm, which also involve the location decision of FDI through MNEs, are based within the host country and include
low costs of labour, agglomeration effects, good institutions that guarantee protection of knowledge, political stability, and infrastructure. As Dunning has noted, these factors are context-specific, giving more credence to country studies. The integration of Business School theories and the traditional trade theories have generated new trade theories which have provided an alternative framework for analyzing FDI and MNE activity. The new trade theories combine both ownership and location advantages with technology and country characteristics and explain both horizontal and vertical FDI. Horizontal FDI, for instance, is explained using the proximity-concentration hypothesis, while vertical FDI is explained using the factor-proportions hypothesis. This area has recently been complemented by the knowledge-capital models that allow for both forms of FDI as special cases. These models have been modified to explain the role of ‘hybrid’ or complex FDI, which is neither purely horizontal nor purely vertical, in addition to third-country effects. An important observation is that all these models explain various aspects of FDI, and although they converge in some aspects, a general theory of FDI is yet to be developed.

In terms of empirical work, the literature on the impact of FDI as well as the determinants of FDI has been studied in both partial and general-equilibrium models. A primary issue with application of the general equilibrium theory to an empirical specification is the complexity of the theoretical models that generally do not have closed solutions. Testing for most of the hypotheses, especially within the North-South interactions, has been difficult because some of the factors are inherently unobservable. As a result, various proxies have been used in the measurement of various variables such as research and development, skills of labour, and imitation and copying. These proxies could lead to wrong conclusions as they are just approximations of the unobserved variables. In addition, the limitation in data as well as theoretical assumptions that cannot be operationalized present a problem for partial equilibrium studies. The concern with evidence from partial equilibrium models is that they ignore important long-term equilibrium factors that may affect FDI decisions and locations. This can lead to omitted variable bias in the empirical specification. This is particularly a concern especially when the studies are based on cross-sectional data. Nevertheless, most of the empirical studies support the argument that FDI and trade promote growth and improve welfare in the South but depend to a large extent on the existence of absorption capacities. However, there is need for more South-South FDI studies, to provide more evidence of the possible contributions of FDI in development in comparative perspective. The latter is the contribution of this study.